

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

**THIS PAGE BLANK (USPTO)**

85759

From:  
Sent:  
To:  
Subject:

*[Handwritten signature]*

Li, Ruixiang  
Tuesday, February 04, 2003 11:23 AM  
STIC-Biotech/ChemLib  
Sequence search of Application NO: 09/775,181

Please do a standard search of SEQ ID NOS: 1, 2, and 4 against both the commercial and interference nucleic acid databases.

Thank you very much!

Ruixiang Li  
GAU 1646  
CM1 8E18  
Mail Box 10C01  
306-0282

*[Large handwritten circled number: 10019]*

Point of Contact:  
Toby Port  
Technical Info. Specialist  
CM1 6A04  
703-308-3534

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 2/5  
Date Completed: 2/20  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2003, 04:20:58 ; Search time 9086 Seconds  
(without alignments)  
11684.684 Million cell updates/sec

Title: US-09-775-181-1  
Perfect score: 3648  
Sequence: 1 atgggagcatggttacc.....gggatagtttaagtgtag 3648

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Maximum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_hgo\_hum.\*
- 40: em\_hgo\_mus.\*
- 41: em\_hgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	3648	100.0	3648	6	AX208076	AX208076 Sequence
2	3525	96.6	7148	6	AX208080	AX208080 Sequence
3	2575	70.6	2933	9	AK095120	AK095120 Homo sapi
4	1758.4	48.2	4789	9	AB032962	AB032962 Homo sapi
5	1756.4	48.1	1800	6	AX208078	AX208078 Sequence
6	1504.4	41.2	130039	9	AL355587	AL355587 Human DNA
c 7	1504.4	41.2	152209	2	AL591023	AL591023 Homo sapi
8	1504.4	41.2	169608	2	AL591477	AL591477 Homo sapi
c 9	1079.6	29.6	153265	2	AC127983	AC127983 Rattus no
10	1077.8	29.5	1857	9	AB052146	AB052146 Macaca fa
11	1026.6	28.1	209802	2	AL772387	AL772387 Mus muscu
12	903	24.8	111117	2	AL358932	AL358932 Homo sapi
13	903	24.8	163680	9	AL139821	AL139821 Human DNA
c 14	736.6	20.2	174215	2	AC094823	AC094823 Rattus no
c 15	730.2	20.0	163199	2	AC109216	AC109216 Mus muscu
16	730.2	20.0	172612	2	AC110523	AC110523 Mus muscu
c 17	601	16.5	2434	9	AK094758	AK094758 Homo sapi
18	583.4	16.0	135024	2	AC106583	AC106583 Rattus no
c 19	227.2	6.2	143878	2	AL359974	AL359974 Homo sapi
20	225.6	6.2	159231	9	AL161654	AL161654 Human DNA
21	200	5.5	153265	2	AC127983	AC127983 Rattus no
22	175	4.8	141031	2	AC109297	AC109297 Mus muscu
23	158.2	4.3	794	9	HS338329	HS338329 Homo sapi
24	121.6	3.3	944	9	HS326137	HS326137 Homo sapi
c 25	119.2	3.3	114538	2	AC115992	AC115992 Homo sapi
26	117.6	3.2	158503	2	AC124789	AC124789 Homo sapi
27	117	3.2	955	9	HS338498	HS338498 Homo sapi
c 28	108	3.0	160390	2	AL445244	AL445244 Homo sapi
c 29	96.4	2.6	233204	10	AL596088	AL596088 Mouse DNA
30	95.4	2.6	163199	2	AC109216	AC109216 Mus muscu
31	93.8	2.6	174215	2	AC094823	AC094823 Rattus no
c 32	93	2.5	183685	2	AC099199	AC099199 Rattus no
33	89.8	2.5	163836	2	AC118154	AC118154 Rattus no
c 34	83.8	2.3	71346	2	AC109295	AC109295 Mus muscu
35	76.4	2.1	183685	2	AC099199	AC099199 Rattus no
36	73.8	2.0	60097	2	AC127484	AC127484 Homo sapi
c 37	73.2	2.0	60097	2	AC127484	AC127484 Homo sapi
38	71	1.9	504	9	HS3323276	HS3323276 Homo sapi
c 39	65.2	1.8	125020	9	AF429315	AF429315 Homo sapi
40	64.2	1.8	125020	9	AF429315	AF429315 Homo sapi
c 41	62	1.7	7218	6	I66494	I66494 Sequence 14
42	60.6	1.7	36583	1	SC5H1	SC5H1 Streptomy
43	59.8	1.6	2010	6	AX254615	AX254615 Sequence
44	57.2	1.6	172838	2	AC128878	AC128878 Rattus no
c 45	57	1.6	141031	2	AC109297	AC109297 Mus muscu

ALIGNMENTS

RESULT 1  
AX208076  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AX208076  
Sequence 1 from Patent WO0157086.  
AX208076  
AX208076.1 GI:15422580  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 3648)  
Donoho,G. and Hilbun,E.  
Novel human g protein coupled receptor proteins and polynucleotides  
encoding the same

linear PAT 31-AUG-2001



[illegible]

Db 489 GCTGTGGCGCCAGCGCGACCCGACCCGAGGAGCGCGATTCCCTCGAGAGAGGACCCCG 548  
Qy 121 AAGGGAAAGCCGACGCGCCAGCAGCGGGTCGAGCCTCTGGCTCGGACTCTCTCGGCTCCC 180  
Db 549 AAGGGAAAGCCGACGCGCCAGCAGCGGGTCGAGCCTCTGGCTCGGACTCTCTCGGCTCCC 608  
Qy 181 TGGAGCGCTCCACCGATGGCACCACTTCTGGCGCAGAAACTCGCCGAGGAGTGCCCATG 240  
Db 609 TGGAGCGCTCCACCGATGGCACCACTTCTGGCGCAGAAACTCGCCGAGGAGTGCCCATG 568  
Qy 241 GAGGTGGCTCTTACCTCTACACCGGGGACTCCACAGCTGAAGCGAGCCAACTGTCTCC 300  
Db 669 GAGGTGGCTCTTACCTCTACACCGGGGACTCCACAGCTGAAGCGAGCCAACTGTCTCC 728  
Qy 301 GCGCGCTACGAGTGGCGGCTGCGGGGAGTGGCCAGCCCTGGCCAGCGCGCACCCC 360  
Db 729 GCGCGCTACGAGTGGCGGCTGCGGGGAGTGGCCAGCCCTGGCCAGCGCGCACCCC 788  
Qy 361 TCCTTGCACCGGGCGTGGACACACTGACACAGCCACCAACTTCTCAACGTGATGCTG 420  
Db 789 TCCTTGCACCGGGCGTGGACACACTGACACAGCCACCAACTTCTCAACGTGATGCTG 848  
Qy 421 CAGAGCAATAAGTTCGGGAGCAGAACTTTCAGGACGACTGGATGGTGTACAGCGCTG 480  
Db 849 CAGAGCAATAAGTTCGGGAGCAGAACTTTCAGGACGACTGGATGGTGTACAGCGCTG 908  
Qy 481 GTGTGGAGCTTCTGGAGGCGAGCCAGCATCTCCGGGCGGCCATCACCTTCAGCACCC 540  
Db 909 GTGTGGAGCTTCTGGAGGCGAGCCAGCATCTCCGGGCGGCCATCACCTTCAGCACCC 968  
Qy 541 GATTGGCTTCGCGCACCGGCCACAGGTCTTCTCCAGGCCACGCGGAGGAGAGCGCG 600  
Db 969 GATTGGCTTCGCGCACCGGCCACAGGTCTTCTCCAGGCCACGCGGAGGAGAGCGCG 1028  
Qy 601 ATCTGTCTCCAAAGACTGTCTCTCCGCAOCCCACTGTGGCAAGCCCACTCTGGAGACC 660  
Db 1029 ATCTGTCTCCAAAGACTGTCTCTCCGCAOCCCACTGTGGCAAGCCCACTCTGGAGACC 1088  
Qy 661 GAGTGTTCACGCGCTCGCGCGCAAGTGGAGGCGCCACTTACCGCGCGCGCCCAAT 720  
Db 1089 GAGTGTTCACGCGCTCGCGCGCAAGTGGAGGCGCCACTTACCGCGCGCGCCCAAT 1148  
Qy 721 CAGGCGCCCGGGGCTCGGCGCACAGCTGTGGCGCGCAAGGAGCGGCTCGGCGGGGACAAG 780  
Db 1149 CAGGCGCCCGGGGCTCGGCGCACAGCTGTGGCGCGCAAGGAGCGGCTCGGCGGGGACAAG 1208  
Qy 781 AGCCACTCAAGTGTCTCCGCTTATCTGGAGTGGGAGACGGGAGTTACAGCCCGGG 840  
Db 1209 AGCCACTCAAGTGTCTCCGCTTATCTGGAGTGGGAGACGGGAGTTACAGCCCGGG 1268  
Qy 841 TGGCTGGTTACTCTTCTCTGCCATCTACGGTTGCGAGCCTTAACCTGGTCCCGGAATTC 900  
Db 1269 TGGCTGGTTACTCTTCTCTGCCATCTACGGTTGCGAGCCTTAACCTGGTCCCGGAATTC 1328  
Qy 901 AGGGGTGCATGAAAGTTGACATAATCTTCAGAAAGTGGACATTGACCAATGTCTCAAGT 960  
Db 1329 AGGGGTGCATGAAAGTTGACATAATCTTCAGAAAGTGGACATTGACCAATGTCTCAAGT 1388  
Qy 961 GATGCTGGTTTTCAGGAACATCAAAATGCCACCTCAACAATTCAGAGTGTATGCCAAT 1020  
Db 1389 GATGCTGGTTTTCAGGAACATCAAAATGCCACCTCAACAATTCAGAGTGTATGCCAAT 1448  
Qy 1021 AAAGCCCTAGGATTCGTTCTGGAGCCTATGAGTGCATTTCGAAAGCAGGATTCATCAT 1080  
Db 1449 AAAGCCCTAGGATTCGTTCTGGAGCCTATGAGTGCATTTCGAAAGCAGGATTCATCAT 1508  
Qy 1081 COTGAGCTTACCAAGTGAACAACCTTCGGGAGAGGGGTCCGGATTCAGCATATTTTCAGGA 1140  
Db 1509 COTGAGCTTACCAAGTGAACAACCTTCGGGAGAGGGGTCCGGATTCAGCATATTTTCAGGA 1568  
Qy 1141 AGTACAAAAGATGTCTCAGAGAGCCCTATGTCTGCCCTACCTTCGAGGAGGCGTGCACC 1200  
Db 1569 AGTACAAAAGATGTCTCAGAGAGCCCTATGTCTGCCCTACCTTCGAGGAGGCGTGCACC 1628

Qy 1201 TTCTGTGCTGATGACAGCCCATGCTTCGTCAGGAAGATAAGTATTTACGACTTGCCATC 1260  
Db 1629 TTCTGTGCTGATGACAGCCCATGCTTCGTCAGGAAGATAAGTATTTACGACTTGCCATC 1688  
Qy 1261 ATCTCTCTTCCAAAGGCTGTGATGCTGCTCGACTTCGTTAGCATGCTGTGCTGTACAC 1320  
Db 1689 ATCTCTCTTCCAAAGGCTGTGATGCTGCTCGACTTCGTTAGCATGCTGTGCTGTACAC 1748  
Qy 1321 TTTCGCAAGCAAGAGATCCGGGCATCGGGCCTTATTCCTGTTGGAACGATCTTTTT 1380  
Db 1749 TTTCGCAAGCAAGAGATCCGGGCATCGGGCCTTATTCCTGTTGGAACGATCTTTTT 1808  
Qy 1381 GGATCTCTGCTCTATACTTCTCCAGTGTGTTTGTACTTTGAGCCAGCACATTTGCG 1440  
Db 1809 GGATCTCTGCTCTATACTTCTCCAGTGTGTTTGTACTTTGAGCCAGCACATTTGCG 1868  
Qy 1441 TGTATTTCTCCTAAGATGGGCTGCTCTTCGCTTTGCTACTGTTTACGGAACTGTCACT 1500  
Db 1869 TGTATTTCTCCTAAGATGGGCTGCTCTTCGCTTTTGTACTGTTTACGGAACTGTCACT 1928  
Qy 1501 CTCAAACTTCACAGGGTTTGAAGTGTTCCTTCACGAACGGCTCAACGAATTCATAT 1560  
Db 1929 CTCAAACTTCACAGGGTTTGAAGTGTTCCTTCACGAACGGCTCAACGAATTCATAT 1988  
Qy 1561 ATGACTGGCGGACGGTTCATGAGGATGCTGGCAGTAATACTCTTGGTAGTGTTTGGTT 1620  
Db 1989 ATGACTGGCGGACGGTTCATGAGGATGCTGGCAGTAATACTCTTGGTAGTGTTTGGTT 2048  
Qy 1621 CTCATTGGCTGGACTTCATCTGTGCCAGAAATTTGGAGAAACAGATTTCACTATTGGC 1680  
Db 2049 CTCATTGGCTGGACTTCATCTGTGCCAGAAATTTGGAGAAACAGATTTCACTATTGGC 2108  
Qy 1681 CAGGGAAACATCCGATCACTCATCTTCAATATGTGCTCATTTGACCGTGGGACTAC 1740  
Db 2109 CAGGGAAACATCCGATCACTCATCTTCAATATGTGCTCATTTGACCGTGGGACTAC 2168  
Qy 1741 ATGACAGCAGTTG----- 1753  
Db 2169 ATGACAGCAGTTGATGATGTTGTTGCTGATGATGTTTACCATTTCCTCAGTAG 2228  
Qy 1754 ----- 1753  
Db 2229 CATTAACACTGACTTCTTTTAAAGTCTTGGCAGACTCTTAGTAAAGCCTAAACTAATCT 2288  
Qy 1754 -----CTGAATTTTATTCCTTGTGGGTGTTTATCTCTGCTATGCTGCGGACAG 1807  
Db 2289 CACCTACTGAATTTTATTCCTTGTGGGTGTTTATCTCTGCTATGCTGCTGCGGACAG 2348  
Qy 1808 TCCCATCGSCATTCATGAGCCCGCTATATGGCTGTTGCGAGTTGCAGTTACAGTCAATCA 1867  
Db 2349 TCCCATCGSCATTCATGAGCCCGCTATATGGCTGTTGCGAGTTGCAGTTACAGTCAATCA 2408  
Qy 1868 TCTCTGCTATATTCATACAATTAGATTGTTTCTTGGCCTCAAGACTTCAGTCTGATTGGA 1927  
Db 2409 TCTCTGCTATATTCATACAATTAGATTGTTTCTTGGCCTCAAGACTTCAGTCTGATTGGA 2468  
Qy 1928 TGTGTAGCTGTATTTTGCACATACACTCATTTGACTGTGTGACAGTCACTATTTGGTGTCT 1987  
Db 2469 TGTGTAGCTGTATTTTGCACATACACTCATTTGACTGTGTGACAGTCACTATTTGGTGTCT 2528  
Qy 1988 TGATTCCAAAGCTTTTCATTTCAAGCAATAACCCAGAGATGATTTCTCAGAAAGCAT 2047  
Db 2529 TGATTCCAAAGCTTTTCATTTCAAGCAATAACCCAGAGATGATTTCTCAGAAAGCAT 2588  
Qy 2048 ATGAGGATGAGTACAGATGGCGGATCTGGATCCTACTGAAACAGCAGTATCAATTCAG 2107  
Db 2589 ATGAGGATGAGTACAGATGGCGGATCTGGATCCTACTGAAACAGCAGTATCAATTCAG 2648  
Qy 2108 CTTGGAGTACAGCAGCTTGGATCCAGAGGACATTCGGGACGAGCTGAAAAAATCTCTATG 2167  
Db 2649 CTTGGAGTACAGCAGCTTGGATCCAGAGGACATTCGGGACGAGCTGAAAAAATCTCTATG 2708

QY 2168 CCCAACTGGAATATATATAAAGAAAGAGATGATCACAAACAAACCCACCTCCAGAAAA 2227  
Db 2709 CCCAACTGGAATATATATAAAGAAAGAGATGATCACAAACAAACCCACCTCCAGAAAA 2768  
QY 2228 ACGGTGTCGGAAGAGGCGCTAGCTCGTTCCATCATGAGAGCGCATTCAGGAGATCCCGAG 2287  
Db 2769 ACGGTGTCGGAAGAGGCGCTAGCTCGTTCCATCATGAGAGCGCATTCAGGAGATCCCGAG 2828  
QY 2288 AGACAGTCAGCGCGGAGTGTCTAAAGAGGCAAGAGGCGCGCCAGCATGCACAGGCA 2347  
Db 2829 AGACAGTCAGCGCGGAGTGTCTAAAGAGGCAAGAGGCGCGCCAGCATGCACAGGCA 2888  
QY 2348 AAGGCACTGCCCTCATCAGGAAGAACCCGCCAGAGTCTTCAGGGAACACAGAGGAATCCA 2407  
Db 2889 AAGGCACTGCCCTCATCAGGAAGAACCCGCCAGAGTCTTCAGGGAACACAGAGGAATCCA 2948  
QY 2408 AGGAGAGACCCCTGAAACCCGAGTCTCTCACTCAAGAAATCCACAGCATTTATGACC 2467  
Db 2949 AGGAGAGACCCCTGAAACCCGAGTCTCTCACTCAAGAAATCCACAGCATTTATGACC 3008  
QY 2468 ACGTGAGAGACCAACCGGAAGAGTCCAGTAGCTTACCCACAGAAAACCAAGAGGAGAGA 2527  
Db 3009 ACGTGAGAGACCAACCGGAAGAGTCCAGTAGCTTACCCACAGAAAACCAAGAGGAGAGA 3068  
QY 2528 CAACAGAAAATTCACACTGGAATCCCTGTGCGGTAAAAAATAACAAAACTAAACAA 2587  
Db 3069 CAACAGAAAATTCACACTGGAATCCCTGTGCGGTAAAAAATAACAAAACTAAACAA 3128  
QY 2588 AAGACAGCGAGGCTGAGTCCACGAGTCCGTGCGCGTGTGTGCAAGTCAGCAACGCGTC 2647  
Db 3129 AAGACAGCGAGGCTGAGTCCACGAGTCCGTGCGCGTGTGTGCAAGTCAGCAACGCGTC 3188  
QY 2648 ACAACCTCAGCTCAGAGAGAAAATCGGCGACCCACAGCAATCGATGTTACAGAACTCTC 2707  
Db 3189 ACAACCTCAGCTCAGAGAGAAAATCGGCGACCCACAGCAATCGATGTTACAGAACTCTC 3248  
QY 2708 TCAGTGTATACAGACGCCCAAGGAGAGACTCTCTGGATTTAGCTGGGAAAACCCAAACAG 2767  
Db 3249 TCAGTGTATACAGACGCCCAAGGAGAGACTCTCTGGATTTAGCTGGGAAAACCCAAACAG 3308  
QY 2768 CAGGTGTGAAACGCACTAAATCCAGAACCTTTCGCAAAAGATTAAGAGACAACA 2827  
Db 3309 CAGGTGTGAAACGCACTAAATCCAGAACCTTTCGCAAAAGATTAAGAGACAACA 3368  
QY 2828 GAAATCACTCAAAATCTGATAACACAGACTAAAGATCTCTGCCCCCAAAACTCAAAATC 2887  
Db 3369 GAAATCACTCAAAATCTGATAACACAGACTAAAGATCTCTGCCCCCAAAACTCAAAATC 3428  
QY 2888 CTGGGAGGAGCCAGAACGCTCAGAAATCTGGGATTTATGAACAACAAGGGGTCAACC 2947  
Db 3429 CTGGGAGGAGCCAGAACGCTCAGAAATCTGGGATTTATGAACAACAAGGGGTCAACC 3488  
QY 2948 CCACCACTGCCAAATCTTGACCTGAACCCAGGCGACCCAGATGAAGGACACTTTGACA 3007  
Db 3489 CCACCACTGCCAAATCTTGACCTGAACCCAGGCGACCCAGATGAAGGACACTTTGACA 3548  
QY 3008 TTGGGAGGTGTCTCTTGGGAGGTTTATGACCTGACCCCTGGTCTGCTGCTTCAGAAAT 3067  
Db 3549 TTGGGAGGTGTCTCTTGGGAGGTTTATGACCTGACCCCTGGTCTGCTGCTTCAGAAAT 3608  
QY 3068 CAAAAGTTCAAAGACAGTATCTATGTGGCTTCGAAATGGAGAAAACCCCACTTTT 3127  
Db 3609 CAAAAGTTCAAAGACAGTATCTATGTGGCTTCGAAATGGAGAAAACCCCACTTTT 3668  
QY 3128 CTTAAAGGAGAAATCTCACCAAGCCCTAAGGCGAGCTGAGGTGTTGTGAGCAATCCAATC 3187  
Db 3669 CTTAAAGGAGAAATCTCACCAAGCCCTAAGGCGAGCTGAGGTGTTGTGAGCAATCCAATC 3728  
QY 3188 AGAAGCGCATAGATGAAGCTGAAGTATGCTTTGGGAGAGCCCAAGCCAGTCCATTTTGG 3247  
Db 3729 AGAAGCGCATAGATGAAGCTGAAGTATGCTTTGGGAGAGCCCAAGCCAGTCCATTTTGG 3788  
QY 3248 AAGATGAGAGGCTTTTGTATTTCCAGACTCCAGTCTCCCGAGAGAGGGCAAAAGAGGAGA 3307

Db 3789 AAGATGAGAGGCTTTTGTATTTCCAAAGACTCCAGTTCTCCCGAGAGGGCAAAAGAGGAGA 3848  
QY 3308 ACGGAGGTTCAGCTCGTCAGCAATGTGTGCTGGCGAGAGCGCAAGAACTGCCCCCA 3367  
Db 3849 ACGGAGGTTCAGCTCGTCAGCAATGTGTGCTGGCGAGAGCGCAAGAACTGCCCCCA 3908  
QY 3368 AAGCTGTAGCTCAAAAACAGAGAAATGAAAATCTCAACAAATAGGACACCAGGAAAAA 3427  
Db 3909 AAGCTGTAGCTCAAAAACAGAGAAATGAAAATCTCAACAAATAGGACACCAGGAAAAA 3968  
QY 3428 AGACATCTTCTTCAGGAGAAATGTGCGTGCCTCTATAACTCAAGTATAACTTTCCAGC 3487  
Db 3969 AGACATCTTCTTCAGGAGAAATGTGCGTGCCTCTATAACTCAAGTATAACTTTCCAGC 4028  
QY 3488 AACCTTTAACATCAGCAGCAGAGGTTTGTCTTGGAGTTTGAGACCCAGCTCAACCA 3547  
Db 4029 AACCTTTAACATCAGCAGCAGAGGTTTGTCTTGGAGTTTGAGACCCAGCTCAACCA 4088  
QY 3548 ATGCTGGAAGAGTGTAGCTTTACCTGCTCTTCTGCTCTTAAGTCAAAATAGATAGCAG 3607  
Db 4089 ATGCTGGAAGAGTGTAGCTTTACCTGCTCTTCTGCTCTTAAGTCAAAATAGATAGCAG 4148  
QY 3608 GGCCTAGGAAGAGAGATCTGGGATAGTTTAAAGTGTAG 3648  
Db 4149 GGCCTAGGAAGAGAGATCTGGGATAGTTTAAAGTGTAG 4189  
RESULT 3  
AK095120 Homo sapiens cdna FLJ37801 fis, clone BRSSN2000312. PRI 15-JUL-2002  
LOCUS  
DEFINITION  
AK095120  
ACCESSION  
AK095120.1 GI:21754315  
VERSION  
oligo capping; fis (full insert sequence).  
KEYWORDS  
Homo sapiens substantia nigra cdna to mRNA, clone\_lib:BRSSN2  
SOURCE  
clone:BRSSN2000312.  
ORGANISM  
Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
AUTHORS  
Kanehori,K., Ishibashi,T., Chiba,Y., Fujimori,K., Hiraoka,S.,  
Tanai,H., Watanabe,S., Ishida,S., Ono,Y., Hotuta,T., Watanabe,M.,  
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,  
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,  
Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,  
Wagatsuma,M., Takahashi-Fujii,A., Oshima,A., Sugiyama,A.,  
Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.  
and Isogai,T.  
NEDO human cdna sequencing project  
TITLE  
JOURNAL  
2 (bases 1 to 2933)  
REFERENCE  
AUTHORS  
Isogai,T. and Yamamoto,J.  
TITLE  
JOURNAL  
Direct Submission  
Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
COMMENT  
NEDO human cdna sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cdna full insert sequencing:  
Research Association for Biotechnology (RAB); cdna library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
FEATURES  
Location/Qualifiers  
1. .2933  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="BRSSN2000312"  
/tissue\_type="substantia nigra"  
/clone\_lib="BRSSN2"  
/note="cloning vector: pME18SFL3"

BASE COUNT	585	a	831	c	770	g	647	t
ORIGIN								
Query Match	70.6%; Score 2575; DB 9; Length 2933;							
Best Local Similarity	100.0%; Pred. No. 0;							
Matches 2575; Conservative	0; Mismatches 0; Indels 0; Gaps							
Qy	1	ATGGAGCCATGGCTTACCCCTTTACTCTCTGCCTCTCTCTTGTCTGCTGCTGAGTGGGA	60					
Db	359	ATGGAGCCATGGCTTACCCCTTTACTCTCTGCCTCTCTCTTGTCTGCTGAGTGGGA	418					
Qy	61	GCTGTTGGGCGAGCGCGACCCCCAAGAGACGGCGGGATTCCCTCTGAGAGAGACCCCG	120					
Db	419	GCTGTTGGGCGACGCGCGACCCCCAAGAGACGGCGGGATTCCCTCTGAGAGAGACCCCG	478					
Qy	121	AAGGGGAAGCGGCAGCGCCAGCAGCGGGTGCAGGCTCTGCTTCGACTCTCTCGGCTCCC	180					
Db	479	AAGGGGAAGCGGCAGCGCCAGCAGCGGGTGCAGGCTCTGCTTCGACTCTCTCGGCTCCC	538					
Db	539	TGGAGCGCTCCACCGATGSCACCATCTTGGCGCAGAACTCGCGAGAGGTGCCATG	598					
Qy	241	GACGTGGCCTTCTTACTCTTACACCGGGGACTTCCCACAGCTGAAGAGAGCAACTGCTCC	300					
Db	599	GACGTGGCCTTCTTACTCTTACACCGGGGACTTCCCACAGCTGAAGAGAGCAACTGCTCC	658					
Qy	301	GGCCGCTACAGATTGGGGCGCTGCGGGGAAGTGGCCAGCGCTTGGCCAGCGCGACCCC	360					
Db	659	GGCCGCTACAGATTGGGGCGCTGCGGGGAAGTGGCCAGCGCTTGGCCAGCGCGACCCC	718					
Qy	361	TCCTTGACCGGGCGCTGGACACACTTGACACAGCCACCACTTCTCAAGGTGATGCTG	420					
Db	719	TCCTTGACCGGGCGCTGGACACACTTGACACAGCCACCACTTCTCAAGGTGATGCTG	778					
Qy	421	CAGAGCAATTAAGTCGGGGAGCAGAACTTGCAGGAGGACCTGGATTGTTACAGCGCTG	480					
Db	779	CAGAGCAATTAAGTCGGGGAGCAGAACTTGCAGGAGGACCTGGATTGTTACAGCGCTG	838					
Qy	481	GTGTGGAGCTTCTTGGAGGGCAGGCCACGATCTCCGGGGGGCCATCACCTTCAGACCC	540					
Db	839	GTGTGGAGCTTCTTGGAGGGCAGGCCACGATCTCCGGGGGGCCATCACCTTCAGACCC	898					
Qy	541	GATTGCTGTCCGACACGGCCCCACAGGTCTTCTCCAGGCCACGCGGAGGAGAGCGCG	600					
Db	899	GATTGCTGTCCGACACGGCCCCACAGGTCTTCTCCAGGCCACGCGGAGGAGAGCGCG	958					
Db	601	ATCTGTCTCAAGACCTGTCTCTCCGCACCCCACTTGGCCACAGCCACTCTGGAGACC	660					
Db	959	ATCTGTCTCAAGACCTGTCTCTCCGCACCCCACTTGGCCACAGCCACTCTGGAGACC	1018					
Qy	661	GAGTGTGTTCCACGGCTCCGGCGCAGTGGAGGCCCACTTACACCGCGCGGCCCAAT	720					
Db	1019	GAGTGTGTTCCACGGCTCCGGCGCAGTGGAGGCCCACTTACACCGCGCGGCCCAAT	1078					
Qy	721	CAGGGCCCCGGGCGCTTGGGCCACAGTGGCGGCCAAGACGGGCTCGCGGGGACAAG	780					
Db	1079	CAGGGCCCCGGGCGCTTGGGCCACAGTGGCGGCCAAGACGGGCTCGCGGGGACAAG	1138					
Qy	781	AGCCACTTCAAGTGTGCTCCGGCTTATCTGGAGTGCAGAGACGGGAGTTACAAGCCCCGG	840					
Db	1139	AGCCACTTCAAGTGTGCTCCGGCTTATCTGGAGTGCAGAGACGGGAGTTACAAGCCCCGG	1198					
Qy	841	TGGCTGTTACTCTTTCCTCTGCCATCTACGGGTTGCAGCCTAACCTGGTCCCGGAATTC	900					
Db	1199	TGGCTGTTACTCTTTCCTCTGCCATCTACGGGTTGCAGCCTAACCTGGTCCCGGAATTC	1258					
Qy	901	AGGGGTCTATGAAGTTGACATAATCTTCAGAAGTGGACATTGACCAATGCTCAAGT	960					
Db	1259	AGGGGTCTATGAAGTTGACATAATCTTCAGAAGTGGACATTGACCAATGCTCAAGT	1318					
Qy	961	GATGGCTGGTTTTTCAGGAACCTCATAAATGCCACCTCAACAATTCAGAGTGTATGCCAATT	1020					



QY	2101	AATT	CAGCCTGGAGT	GACACAGCTT	GGATCC	CAGAGN	CATT	CGGG	ACCAGCT	GA	AAAA	2160
Db	2459	AATT <td>CAGCCTGGAGT <th>GACACAGCTT</th> <th>GGATCC</th> <th>CAGAGN</th> <th>CATT</th> <th>CGGG</th> <th>ACCAGCT</th> <th>GA</th> <th>AAAA</th> <th>2518</th> </td>	CAGCCTGGAGT <th>GACACAGCTT</th> <th>GGATCC</th> <th>CAGAGN</th> <th>CATT</th> <th>CGGG</th> <th>ACCAGCT</th> <th>GA</th> <th>AAAA</th> <th>2518</th>	GACACAGCTT	GGATCC	CAGAGN	CATT	CGGG	ACCAGCT	GA	AAAA	2518
QY	2161	CTCT <td>TATGCCCAACT <th>TGGAAATAT</th> <th>TATATA</th> <th>AAAA</th> <th>GAAGAAG</th> <th>TATGAT</th> <th>CACAAAC</th> <th>CAACCC</th> <th>CCACCTC</th> <th>2220</th> </td>	TATGCCCAACT <th>TGGAAATAT</th> <th>TATATA</th> <th>AAAA</th> <th>GAAGAAG</th> <th>TATGAT</th> <th>CACAAAC</th> <th>CAACCC</th> <th>CCACCTC</th> <th>2220</th>	TGGAAATAT	TATATA	AAAA	GAAGAAG	TATGAT	CACAAAC	CAACCC	CCACCTC	2220
Db	2519	CTCT <td>TATGCCCAACT <th>TGGAAATAT</th> <th>TATATA</th> <th>AAAA</th> <th>GAAGAAG</th> <th>TATGAT</th> <th>CACAAAC</th> <th>CAACCC</th> <th>CCACCTC</th> <th>2578</th> </td>	TATGCCCAACT <th>TGGAAATAT</th> <th>TATATA</th> <th>AAAA</th> <th>GAAGAAG</th> <th>TATGAT</th> <th>CACAAAC</th> <th>CAACCC</th> <th>CCACCTC</th> <th>2578</th>	TGGAAATAT	TATATA	AAAA	GAAGAAG	TATGAT	CACAAAC	CAACCC	CCACCTC	2578
QY	2221	CAGAAA <th>AGCGGTGCT</th> <th>CGAAGA</th> <th>GGGCTT</th> <th>AGGT</th> <th>CGTTCC</th> <th>ATCAT</th> <th>GAGACG</th> <th>CATT</th> <th>TACGGAG</th> <th>2280</th>	AGCGGTGCT	CGAAGA	GGGCTT	AGGT	CGTTCC	ATCAT	GAGACG	CATT	TACGGAG	2280
Db	2579	CAGAAA <th>AGCGGTGCT</th> <th>CGAAGA</th> <th>GGGCTT</th> <th>AGGT</th> <th>CGTTCC</th> <th>ATCAT</th> <th>GAGACG</th> <th>CATT</th> <th>TACGGAG</th> <th>2638</th>	AGCGGTGCT	CGAAGA	GGGCTT	AGGT	CGTTCC	ATCAT	GAGACG	CATT	TACGGAG	2638
QY	2281	ATCC <th>CAGACAGT</th> <th>CAAGCC</th> <th>GCAGT</th> <th>GTCT</th> <th>TAAG</th> <th>AGAGAC</th> <th>AGAGGG</th> <th>CGCG</th> <th>CGCATG</th> <th>2340</th>	CAGACAGT	CAAGCC	GCAGT	GTCT	TAAG	AGAGAC	AGAGGG	CGCG	CGCATG	2340
Db	2639	ATCC <th>CAGACAGT</th> <th>CAAGCC</th> <th>GCAGT</th> <th>GTCT</th> <th>TAAG</th> <th>AGAGAC</th> <th>AGAGGG</th> <th>CGCG</th> <th>CGCATG</th> <th>2698</th>	CAGACAGT	CAAGCC	GCAGT	GTCT	TAAG	AGAGAC	AGAGGG	CGCG	CGCATG	2698
QY	2341	ACAGC <th>CAAGGCACT</th> <th>GCCTCAT</th> <th>CATCAG</th> <th>GAAGA</th> <th>AAACCC</th> <th>CCAGAGT</th> <th>CTT</th> <th>CAGG</th> <th>GAACAC</th> <th>2400</th>	CAAGGCACT	GCCTCAT	CATCAG	GAAGA	AAACCC	CCAGAGT	CTT	CAGG	GAACAC	2400
Db	2699	ACAGC <th>CAAGGCACT</th> <th>GCCTCAT</th> <th>CATCAG</th> <th>GAAGA</th> <th>AAACCC</th> <th>CCAGAGT</th> <th>CTT</th> <th>CAGG</th> <th>GAACAC</th> <th>2758</th>	CAAGGCACT	GCCTCAT	CATCAG	GAAGA	AAACCC	CCAGAGT	CTT	CAGG	GAACAC	2758
	2401	AAAT <th>CCAAGGAG</th> <th>AGACCC</th> <th>CTGAAA</th> <th>AAACCG</th> <th>AGTCTT</th> <th>CTC</th> <th>ACTCA</th> <th>AGAA</th> <th>ATCC</th> <th>2460</th>	CCAAGGAG	AGACCC	CTGAAA	AAACCG	AGTCTT	CTC	ACTCA	AGAA	ATCC	2460
Db	2759	AAAT <th>CCAAGGAG</th> <th>AGACCC</th> <th>CTGAAA</th> <th>AAACCG</th> <th>AGTCTT</th> <th>CTC</th> <th>ACTCA</th> <th>AGAA</th> <th>ATCC</th> <th>2818</th>	CCAAGGAG	AGACCC	CTGAAA	AAACCG	AGTCTT	CTC	ACTCA	AGAA	ATCC	2818
QY	2461	TATC <th>CACACGTG</th> <th>AGACAC</th> <th>CAACCG</th> <th>AGAGT</th> <th>CTC</th> <th>AGT</th> <th>AGCTT</th> <th>ACCC</th> <th>CAGAA</th> <th>2520</th>	CACACGTG	AGACAC	CAACCG	AGAGT	CTC	AGT	AGCTT	ACCC	CAGAA	2520
Db	2819	TATC <th>CACACGTG</th> <th>AGACAC</th> <th>CAACCG</th> <th>AGAGT</th> <th>CTC</th> <th>AGT</th> <th>AGCTT</th> <th>ACCC</th> <th>CAGAA</th> <th>2878</th>	CACACGTG	AGACAC	CAACCG	AGAGT	CTC	AGT	AGCTT	ACCC	CAGAA	2878
QY	2521	GAGG <th>GAGACAC</th> <th>AGAAAA</th> <th>ATCC</th> <th>CACACT</th> <th>GGAA</th> <th>TCCTG</th> <th>CGGT</th> <th>TA</th> <th>AAAA</th> <th>2575</th>	GAGACAC	AGAAAA	ATCC	CACACT	GGAA	TCCTG	CGGT	TA	AAAA	2575
Db	2879	GAGG <th>GAGACAC</th> <th>AGAAAA</th> <th>ATCC</th> <th>CACACT</th> <th>GGAA</th> <th>TCCTG</th> <th>CGGT</th> <th>TA</th> <th>AAAA</th> <th>2933</th>	GAGACAC	AGAAAA	ATCC	CACACT	GGAA	TCCTG	CGGT	TA	AAAA	2933

RESULT 4

AB032962

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

AB032962

Homo sapiens mRNA for KIAA1136 protein, partial cds.

AB032962.2

GI:20521775

Homo sapiens adult male brain cDNA to mRNA, clone\_lib:pBluescriptII SK plus clone:hj01467.

Homo sapiens

Mammalia; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Metazoa; Primates; Catarrhini; Hominiidae; Homo.

Hirosawa, M., Nagase, T., Ishikawa, K., Kikuno, R., Nomura, N. and Ohara, O.

Characterization of cDNA clones selected by the GeneMark analysis from size-fractionated cDNA libraries from human brain

DNA Res. 6 (5), 329-336 (1999)

10574461

2 (bases 1 to 4789)

Ohara, O., Nagase, T. and Kikuno, R.

Direct Submission

Submitted (04-OCT-1999)

Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdhainfo@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

On May 9, 20

gene	a representative cDNA sequence for KIAA1136."									
CDS	1..4789	/gene="KIAA1136"	/product="KIAA1136 protein"	48.2%;	Score 1758.4;	DB 9;	Length 4789;			
	<47..1837	/protein_id="BAAB6450.2"	/codon_start=1	Best Local Similarity 99.9%;	Pred. No. 0;					
	/gene="KIAA1136"	/db_xref="GI:205211776"	/note="Start codon is not identified."	Matches 1759;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;		
	/codon_start=1									
	/product="KIAA1136 protein"									
	/protein_id="BAAB6450.2"									
	/db_xref="GI:205211776"									
	/translation="SWIVMSNNPRDIIATEAYEDELDMGRSGYVLSNSAWSSEHSLDIPEDIRDELAKL									
	IPFESSHSSNNPRDIIATEAYEDELDMGRSGYVLSNSAWSSEHSLDIPEDIRDELAKL									
	YAOLEIYKKKMITNPNHLOKRCRKKGLGRSMRRIETETPVYSRQCKSEKDEGADH									
	GTAKGATLIRKPESNGTQKSEETLRNFVSLKRSHTSYDHWIDTEESSSLPTRE									
	SOBEETENTSLTSEKGLTKQKLEDESAESTYPLVCKSAHNLSEKKTHPR									
	TSMLQSLSVIASAKGTGLGAKGTQAGVEERTKSKPLPKDKETNRHNSINDTET									
	KDPAPONSAPPEAPKPKQSGIMKQORVNPNTANSDLNPGTQTKMKNFDIGEYCPWEV									
	YDLTPGPVSESKVQHVSIASEMEKNPTEFSLKEKSHKPKAAEVCOOSNOKRIDKA									
	EYCLVMSQGSILDEKLLISKTPVLPRAKEENGQOPRAANYCAQSGSEELPPKAVAS									
	KTENENLNIQIHGEKHTTSSEENRWYSNNSNFPQOPLTSRAEVCPEWFEETPAQPNAG									
	RSVLPASSALSANKTAGPRKEIWPSEFKV"									
BASE COUNT	1534 a	961 c	965 g	1329 t						
ORIGIN										
Query Match	48.2%;	Score 1758.4;	DB 9;	Length 4789;						
Best Local Similarity	99.9%;	Pred. No. 0;								
Matches 1759;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;						
QY	1889	TTAGATTGTTCTTGCCTCAAGACTTCAGTCTGATTGGATGCTGATGCTGATTTGCAC	1948							
Db	78	TCAGATTGTTCTTGCCTCAAGACTTCAGTCTGATTGGATGCTGATGCTGATTTGCAC	137							
QY	1949	ATACTCATTTGACGTGCACAGTCCACCATTTGGTGTGCTTTTGATTCCTCAAGTTCACATT	2008							
Db	138	ATACTCATTTGACGTGCACAGTCCACCATTTGGTGTGCTTTTGATTCCTCAAGTTCACATT	197							
QY	2009	CAAGCAATAACCCACGAGATGATATTGCTACAGAAGCATATGAGGATGAGCTAGACATGG	2058							
Db	198	CAAGCAATAACCCACGAGATGATATTGCTACAGAAGCATATGAGGATGAGCTAGACATGG	257							
QY	2069	GCCGATCTGGATCTACTCTGAACAGCAGTATCAATTTCAGCCTGGAGTCGACACAGCTGG	2128							
Db	258	GCCGATCTGGATCTACTCTGAACAGCAGTATCAATTTCAGCCTGGAGTCGACACAGCTGG	317							
QY	2129	ATCCAGAGGACATTTCGGGACGAGCTGAAAAACCTCTATGCCCACTGGAATATATAAA	2188							
Db	318	ATCCAGAGGACATTTCGGGACGAGCTGAAAAACCTCTATGCCCACTGGAATATATAAA	377							
QY	2189	GAAGAAGATGATCACAAACACCCCACTCCAGAAAAAGCGTGTCTCGAAGAAGGGCC	2248							
Db	378	GAAGAAGATGATCACAAACACCCCACTCCAGAAAAAGCGTGTCTCGAAGAAGGGCC	437							
QY	2249	TAGTCGTTCCATCATGAGAGCGATTACGGAGATCCCGAGACAGTACGCGGCGAGTGCT	2308							
Db	438	TAGTCGTTCCATCATGAGAGCGATTACGGAGATCCCGAGACAGTACGCGGCGAGTGCT	497							
QY	2309	CTAAAGAGGACAAAGAGGGCGCCACCATGSCACAGCCAAAGGCACCTGCCCTCATCAGGA	2368							
Db	498	CTAAAGAGGACAAAGAGGGCGCCACCATGSCACAGCCAAAGGCACCTGCCCTCATCAGGA	557							
QY	2369	AGAACCCCCCAGAGTCTTCAGGGAAACACAGGGAAATCCAAAGGAGGAGACCCCTGAAAAACC	2428							
Db	558	AGAACCCCCCAGAGTCTTCAGGGAAACACAGGGAAATCCAAAGGAGGAGACCCCTGAAAAACC	617							
QY	2429	GAGTCCTTCTCACTCAAGAAATCCCAAGCAGCTTTATGACACAGGTGAGAGACCAACACGGAAG	2488							
Db	618	GAGTCCTTCTCACTCAAGAAATCCCAAGCAGCTTTATGACACAGGTGAGAGACCAACACGGAAG	677							
QY	2489	AGTCCAGTAGCTACCCACAGAAAGCCAAAGGAGGAGACACACAGAAATTCACACATGG	2548							
Db	678	AGTCCAGTAGCTTACCCACAGAAAGCCAAAGGAGGAGACACACAGAAATTCACACATGG	737							
QY	2549	AATCCCTGTGGGTAAAAAACCTAACACAAAACCTATAAGAGAGACAGCGAGCTCAGTCCA	2608							









```
Db 85828 ATTATGAAACAAAGAGGTCAACCCCACTGCCAATCTTGACCTGAACCCAGCACC 85887
|||||
QY 2983 ACCAGATGAAGGACAACTTTGACATTTGGGAGGTGTGCTTGGGAGGTTTATGACCTG 3042
|||||
Db 85888 ACCAGATGAAGGACAACTTTGACATTTGGGAGGTGTGCTTGGGAGGTTTATGACCTG 85947
|||||
QY 3043 ACCCTGGTCTGTGCTTTCAGAACTAAAGTTCAAAGCAGCTATCTATTGTGGCTTCT 3102
|||||
Db 85948 ACCCTGGTCTGTGCTTTCAGAACTAAAGTTCAAAGCAGCTATCTATTGTGGCTTCT 86007
|||||
QY 3103 GAAATGGAGAAACCCCACTTTTCCCTTAAAGGAGAAATCTCACCACAGCCTTAGGCA 3162
|||||
Db 86008 GAAATGGAGAAACCCCACTTTTCCCTTAAAGGAGAAATCTCACCACAGCCTTAGGCA 86067
|||||
QY 3163 GCTGAGTTTGTGACCAATCCCAATCAGAGCGCATAGATAAGCTCAAGTATGCTCTTGG 3222
|||||
Db 86068 GCTGAGTTTGTGACCAATCCCAATCAGAGCGCATAGATAAGCTCAAGTATGCTCTTGG 86127
|||||
QY 3223 GAGAGCAAGGCCAGTCCATTTTGGAGATGAGAGCTTTTGATTCCAGACTCCAGTT 3282
|||||
Db 86128 GAGAGCAAGGCCAGTCCATTTTGGAGATGAGAGCTTTTGATTCCAGACTCCAGTT 86187
|||||
QY 3283 CTCCAGAGAGGCGCAAGAGGAGACGAGTACGCTCGTGCGCCCAATGTGTGCT 3342
|||||
Db 86188 CTCCAGAGAGGCGCAAGAGGAGACGAGTACGCTCGTGCGCCCAATGTGTGCT 86247
|||||
QY 3343 GGCAGAGCGAAGAACTGCCCCCAAGCTGTAGCATCAAAACAGAGAAATGAATCTC 3402
|||||
Db 86248 GGCAGAGCGAAGAACTGCCCCCAAGCTGTAGCATCAAAACAGAGAAATGAATCTC 86307
|||||
QY 3403 AACCAATAGGACACAGGAGAAAGACATCTTCTCTGAGGAGAAATGCGTGCTCC 3462
|||||
Db 86308 AACCAATAGGACACAGGAGAAAGACATCTTCTCTGAGGAGAAATGCGTGCTCC 86367
|||||
QY 3463 TATACTCAAGTAATTAACCTTCCAGCAACCTTTAATCATCAGCAGAGGTTTGTCTTGG 3522
|||||
Db 86368 TATACTCAAGTAATTAACCTTCCAGCAACCTTTAATCATCAGCAGAGGTTTGTCTTGG 86427
|||||
QY 3523 GAGTTTGAAGCCCAAGCTCAACCAATGCTGGAAGAGTGTAGCTTTACTGCTCTTCT 3582
|||||
Db 86428 GAGTTTGAAGCCCAAGCTCAACCAATGCTGGAAGAGTGTAGCTTTACTGCTCTTCT 86487
|||||
QY 3583 GCTCTAAGTGCAAAATAGATAGCAGGCGCTAGGAAGAGAGATCTGGGATAGTTTAA 3642
|||||
Db 86488 GCTCTAAGTGCAAAATAGATAGCAGGCGCTAGGAAGAGAGATCTGGGATAGTTTAA 86547
|||||
QY 3643 GTGTAG 3648
|||||
Db 86548 GTGTAG 86553
|||||
```

```
RESULT 7
AL591023/3 AL591023 152209 bp DNA linear HTG 19-DEC-2001
LOCUS Homo sapiens chromosome 10 clone RP13-18101, *** SEQUENCING IN
DEFINITION PROGRESS ***, 10 unordered pieces.
ACCESSION AL591023
VERSION AL591023.3 GI:17973954
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE Burton,J.
AUTHORS Direct Submission
TITLE Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Dec 20, 2001 this sequence version replaced gi:13990263.
----- Genome Center
Center: Wellcome Trust Sanger Institute
```

```
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba18101
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 147964 bases at least Q40
Consensus quality: 149796 bases at least Q30
Consensus quality: 150620 bases at least Q20
Insert size: 151309; sum-of-contigs
Quality coverage: 4.61x in Q20 bases; agarose-fp
coverage: 4.08x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 4881: contig of 4881 bp in length
* 4882 4981: gap of 100 bp
* 4982 13992: contig of 9011 bp in length
* 13993 14092: gap of 100 bp
* 14093 59903: contig of 45811 bp in length
* 59904 60003: gap of 100 bp
* 60004 68910: contig of 8907 bp in length
* 68911 69010: gap of 100 bp
* 69011 82621: contig of 13611 bp in length
* 82622 82721: gap of 100 bp
* 82722 92411: contig of 9690 bp in length
* 92412 92511: gap of 100 bp
* 92512 107868: contig of 15357 bp in length
* 107869 107968: gap of 100 bp
* 107969 129485: contig of 21517 bp in length
* 129486 129585: gap of 100 bp
* 129586 138573: contig of 8988 bp in length
* 138574 138673: gap of 100 bp
* 138674 152209: contig of 13536 bp in length.
FEATURES
Location/Qualifiers
1. 152209
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP13-18101"
/clone.lib="RPC1-13.1"
misc_feature 1..4881
/note="assembly_fragment:02296
fragment_chain:1
clone_end:SP6
vector_side:left"
misc_feature 4882..13992
/note="assembly_fragment:00809
fragment_chain:1"
misc_feature 14093..59903
/note="assembly_fragment:01674
fragment_chain:2"
misc_feature 60004..68910
/note="assembly_fragment:01032
fragment_chain:2"
misc_feature 69011..82621
/note="assembly_fragment:00070
fragment_chain:2"
misc_feature 82722..92411
/note="assembly_fragment:01537
fragment_chain:2"
misc_feature 92512..107868
/note="assembly_fragment:00515
```

misc_feature		fragment_chain:2"	
107969..129485		/note="assembly_fragment:00256	
misc_feature		fragment_chain:2"	
129586..138573		/note="assembly_fragment:01678	
misc_feature		fragment_chain:2"	
138674..152209		/note="assembly_fragment:00292	
		clone_end:T7	
vector_side:right"			
BASE COUNT	43984 a 31361 c 29426 g 46534 t	904 others	
ORIGIN			
Query Match 41.2%; Score 1504.4; DB 2; Length 152209;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 1505; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
QY	2143	CGGGACGAGCTGAAAAAAGCTCTATGCCCACTGGAATATATAAAGAAAGAGATGATC	2202
Db	150338	CAGGACGAGCTGAAAAAAGCTCTATGCCCACTGGAATATATAAAGAAAGAGATGATC	150279
QY	2203	ACAAACAACCCCACTCCAGAAAAAGCGGTGCTCGAAGAAGGCGCTAGTCTGTTCCATC	2262
Db	150278	ACAAACAACCCCACTCCAGAAAAAGCGGTGCTCGAAGAAGGCGCTAGTCTGTTCCATC	150219
QY	2263	ATGAGACGATTACGGAGATCCAGAGACAGTCAGCGCGCAGTGTCTAAAGAGGACAG	2322
Db	150218	ATGAGACGATTACGGAGATCCAGAGACAGTCAGCGCGCAGTGTCTAAAGAGGACAG	150159
QY	2323	GAGGGCGCGACCATGGCAGACGACCAAGGCACCTGCCCTCATCAGGAAGAACCCCCAGAG	2382
Db	150158	GAGGGCGCGACCATGGCAGACGACCAAGGCACCTGCCCTCATCAGGAAGAACCCCCAGAG	150099
QY	2383	TCTTCAGGAAACACAGGGAAATCCAAAGGAGGAGACCCCTGAAACCGAGTCTTCATCT	2442
Db	150098	TCTTCAGGAAACACAGGGAAATCCAAAGGAGGAGACCCCTGAAACCGAGTCTTCATCT	150039
QY	2443	AGAAATCCACAGACACTTATGACCACGTGAGAGACCAACGAGTCCAGTACCTA	2502
Db	150038	AGAAATCCACAGACACTTATGACCACGTGAGAGACCAACGAGTCCAGTACCTA	149979
QY	2503	CCACAGAAAGCCAAAGGAGGAGACCAACAGAAAAATTCACACACTGGAATCCCTGTCGGT	2562
Db	149978	CCACAGAAAGCCAAAGGAGGAGACCAACAGAAAAATTCACACACTGGAATCCCTGTCGGT	149919
QY	2563	AAAAAACTAACAAAAAATTAAGAGACAGCGAGGCTGAGTCCACGGAGTCGGTCCG	2622
Db	149918	AAAAAACTAACAAAAAATTAAGAGAGACAGCGAGGCTGAGTCCACGGAGTCGGTCCG	149859
QY	2623	TTGGTGTGCAAGTCAGCAAGCGCTCACAACTCAGCTCAGAGAAGAAAACTGGGCACCA	2682
Db	149858	TTGGTGTGCAAGTCAGCAAGCGCTCACAACTCAGCTCAGAGAAGAAAACTGGGCACCA	149799
QY	2683	CGAATCATGATGTTACAGAAGTCTCTCAGTGTCTATAGCAAGGCCCAAGGAGAACTCTT	2742
Db	149798	CGAATCATGATGTTACAGAAGTCTCTCAGTGTCTATAGCAAGGCCCAAGGAGAACTCTT	149739
QY	2743	GGATTAGCTGGGAAACCCCAACAGACAGTGTGGAAGACGCACTTAATCCAGAAACCT	2802
Db	149738	GGATTAGCTGGGAAACCCCAACAGACAGTGTGGAAGACGCACTTAATCCAGAAACCT	149679
QY	2803	TTGCCAAAAGATTAAGAGACAAACAGAAATCACTCAAAATTCGTATAACACAGACTAAA	2862
Db	149678	TTGCCAAAAGATTAAGAGACAAACAGAAATCACTCAAAATTCGTATAACACAGACTAAA	149619
QY	2863	GATCCTGCCCCCAAACTCAAAATCTCGGGAGGAGGCCCAAGAAAGCCTCAGAAATCTGGG	2922
Db	149618	GATCCTGCCCCCAAACTCAAAATCTCGGGAGGAGGCCCAAGAAAGCCTCAGAAATCTGGG	149559
QY	2923	ATTATGAACACAAAGGGTCAACCCACCCTGCAATTTGACCTGAACCCAGGCACC	2982
Db	149558	ATTATGAACACAAAGGGTCAACCCACCCTGCAATTTGACCTGAACCCAGGCACC	149499
QY	2983	ACCCAGATGAAGGACAACCTTTGACATTTGGGAGGTGTGTCTTGGAGGTTTATGACCTG	3042
Db	149498	ACCCAGATGAAGGACAACCTTTGACATTTGGGAGGTGTGTCTTGGAGGTTTATGACCTG	149439
QY	3043	ACCCCTGGTCTCTGCTTTCAGAAATCAAAAGTTCAAAAGCAGCTATCTATTTGGCTTCT	3102
Db	149438	ACCCCTGGTCTCTGCTTTCAGAAATCAAAAGTTCAAAAGCAGCTATCTATTTGGCTTCT	149379
QY	3103	GAAATGGAGAAAAACCCCACTTTTCCCTTAAAGGAGAAATCTCACCACAAGCCTAAGGCA	3162
Db	149378	GAAATGGAGAAAAACCCCACTTTTCCCTTAAAGGAGAAATCTCACCACAAGCCTAAGGCA	149319
QY	3163	GCTGAGGTTTGTTCAGCAATCCAAATCAGAGCGCATAGATAAGGCTGAAGTATGCCCTTGG	3222
Db	149318	GCTGAGGTTTGTTCAGCAATCCAAATCAGAGCGCATAGATAAGGCTGAAGTATGCCCTTGG	149259
QY	3223	GAGAGCCAAGGCCAGTCCATTTTGGGAAGATGAGAAGCTTTTGAATTTCCAAAGACTCCAGTT	3282
Db	149258	GAGAGCCAAGGCCAGTCCATTTTGGGAAGATGAGAAGCTTTTGAATTTCCAAAGACTCCAGTT	149199
QY	3283	CTCCAGAGAGGGCAAAAGAGGAGACGAGTCAAGCTCTGTCAGCCCAATGTGTGTCT	3342
Db	149198	CTCCAGAGAGGGCAAAAGAGGAGACGAGTCAAGCTCTGTCAGCCCAATGTGTGTCT	149139
QY	3343	GGCAGAGCGCAAGAACTGCCCCCAAGCTGTAGCATCAAAAACAGAGAATCAAAATCTC	3402
Db	149138	GGCAGAGCGCAAGAACTGCCCCCAAGCTGTAGCATCAAAAACAGAGAATCAAAATCTC	149079
QY	3403	AACCAATAGGACACACAGGAAAAAAGACATCTTCTTCTGAGGAGAAATGTGCGTGGCTC	3462
Db	149078	AACCAATAGGACACACAGGAAAAAAGACATCTTCTTCTGAGGAGAAATGTGCGTGGCTC	149019
QY	3463	TATAACTCAAGTAATACTTCCAGCAACTTTAAACATCACGAGCAGAGTTTGTCTCTGG	3522
Db	149018	TATAACTCAAGTAATACTTCCAGCAACTTTAAACATCACGAGCAGAGTTTGTCTCTGG	148959
QY	3523	GAGTTTGAGACCCAGCTCAACCAATGCTGGAAGAAGTGTAGCTTTACCTGCCCTCTCT	3582
Db	148958	GAGTTTGAGACCCAGCTCAACCAATGCTGGAAGAAGTGTAGCTTTACCTGCCCTCTCT	148899
QY	3583	GCTCTAAGTGCATAATAGATAGCAGGGCTAGGAAGAAGAGATCTGGGATAGTTTTAAA	3642
Db	148898	GCTCTAAGTGCATAATAGATAGCAGGGCTAGGAAGAAGAGATCTGGGATAGTTTTAAA	148839
QY	3643	GTGTAG 3648	
Db	148838	GTGTAG 148833	
RESULT 8			
AL591477			
LOCUS			
DEFINITION Homo sapiens chromosome 10 clone RP13-112H19, *** SEQUENCING IN			
PROGRESS ***, 3 unordered pieces.			
ACCESSION AL591477			
VERSION AL591477.2 GI:17973980			
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
AUTHORS Burton, J.			
TITLE Direct Submission			
JOURNAL Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,			
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:			
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk			
On Dec 20, 2001 this sequence version replaced gi:14141526.			
----- Genome Center			
Center: Wellcome Trust Sanger Institute			
Center code: SC			
Web site: http://www.sanger.ac.uk			



```
Db 144604 GAGTTTGAGACCCAGCTCAACCAATGCTGGAGAAGTGTAGCTTTTACCTGCCCTCTTCT 144663
QY 3583 GCTCTAAGTGCACAAATAGATAGACAGGGCTAGCAAGAGAGATCGGATAGTTTTTAAA 3642
|||||
Db 144664 GCTCTAAGTGCACAAATAGATAGACAGGGCTAGCAAGAGAGATCGGATAGTTTTTAAA 144723
QY 3643 GTGTAG 3648
|||||
Db 144724 GTGTAG 144729

RESULT 9
AC127983/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-413023, *** SEQUENCING IN PROGRESS
ACCESSION AC127983
VERSION AC127983.1 GI:21908544
SYNOPSIS HTG: HTGS-PHASE1.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 153265)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alzbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbacia,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brivka,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Petters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,C., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
REFERENCE
2 (bases 1 to 153265)
Worley,K.C.
Direct Submission
TITLE
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
```

## COMMENT

```
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KAED
Center clone name: CH230-413023
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 101614 bases at least Q40
Consensus quality: 108377 bases at least Q30
Consensus quality: 112616 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1193: contig of 1193 bp in length
1293: gap of unknown length
1294: contig of 1754 bp in length
3047: gap of unknown length
3147: gap of unknown length
4298: contig of 1151 bp in length
4398: gap of unknown length
4299: gap of unknown length
4399: contig of 1234 bp in length
5632: gap of unknown length
5633: contig of 1257 bp in length
6983: gap of unknown length
7083: gap of unknown length
8189: contig of 1100 bp in length
8289: gap of unknown length
10184: contig of 1895 bp in length
10284: gap of unknown length
11465: contig of 1181 bp in length
11565: gap of unknown length
12686: contig of 1121 bp in length
12687: gap of unknown length
14070: contig of 1284 bp in length
14170: gap of unknown length
16004: contig of 1834 bp in length
16104: gap of unknown length
17491: contig of 1387 bp in length
17591: gap of unknown length
19156: contig of 1565 bp in length
19256: gap of unknown length
20499: contig of 1243 bp in length
20599: gap of unknown length
22157: contig of 1558 bp in length
22257: gap of unknown length
25261: contig of 3004 bp in length
25361: gap of unknown length
27933: contig of 2552 bp in length
28013: gap of unknown length
29643: contig of 1630 bp in length
29743: gap of unknown length
31094: contig of 1351 bp in length
31194: gap of unknown length
32639: contig of 1445 bp in length
32739: gap of unknown length
34418: contig of 1679 bp in length
34518: gap of unknown length
36726: contig of 2208 bp in length
36826: gap of unknown length
39441: contig of 2615 bp in length
39541: gap of unknown length
41539: contig of 1998 bp in length
```





```
QY 3402 CAACAAATAGGACACAGAGAAAGAGACATCTTCTTGAGGAGAAATGCGTGGCTC 3461
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 125691 CAACAAATGGAGACAGGAGAGACAGACATCTTCTTGATATCATCTCGCTC 125632

QY 3462 CTATACCTCAAGTAATACTTCAGCAACCTTTAATACATCAGCAGCAGAGTTGTCCTTG 3521
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 125631 CTGTATCTCGAGTAATACTTCCTCAAGCTTAAGCTCAGCAGCGAAGTGTGCTCCG 125572

QY 3522 GGAGTTTGAGACCCAGCTCACCAATCTCGAAGAAGTGTAGCTTACCTGCCCTTC 3581
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 125571 GGAGTTGAGCCCTAGAACCACTGCTGAAGAATCTGTAAGTCTTACCGCCCTTC 125512

QY 3582 TGCTCTAAGTGAATAGATCAGGGCTAGGAAAGAGAGATCTGGGATAGTTTAA 3641
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 125511 TGCTTTAAGTGAAGTAAATACCGGGCCAGGAATAGGAGTCCAGACATTTCAA 125452

QY 3642 AGTGTA 3647
|||||
Db 125451 AATGTA 125446

RESULT 10
AB052146
LOCUS Macaca fascicularis brain cDNA, linear PRI 09-DEC-2000
ACCESSION AB052146
VERSION AB052146.1 GI:11611572
KEYWORDS fis (full insert sequence).
SOURCE Macaca fascicularis adult male cerebellum cortex cDNA to mRNA,
clone_lib:macaque brain cDNA library QcCE clone:QcCE-20671.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Macaca.
1 (sites)
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished
2 (bases 1 to 1857)
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (05-DEC-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
Lab host: TOPI0
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGATGTC)
R. Site2: DraIII (CACCATGTC)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATGTGGCTTTTATTTTATTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al. (University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
( 5' end primer [CTTCTGCTCTAAAGAGCTGG];
3' end primer [CGACCTGAGCTCAGACACA] ).
Location/Qualifiers
1. .1857
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="QcCE-20671"
/sex="male"
/tissue_type="cerebellum cortex"
/clone_lib="macaque brain cDNA library QcCE"
/dev_stage="adult"
213. .1166
CDS
```

```
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB19001.1"
/db_xref="GI:11611573"
/translation="MLQKSLSVIASAKEKTLGLAGKTOTAGVBESAKSQKPLPKDKET
NRHNSDNTETKPTQNSNPABEPRKQSGIMKQORVNPNTTANSLNFGATQMKD
NFDIGEYCPWEYDLTPGPVPSSEKVKHVSIVASEMEKNPTLSLKEKSHHKPAAEV
CQOQNKCTDKAEVCLMGIOGQSILEDEKHFISKTPVLQERAKRENGOPHAAKVCAG
QSEBPPKAVASKTENENLNIQHGEKKTSEBNVRGSSYNSNFPQQLMSRAEVCPPW
EFETPAQPNAGRSVALPVSALSASKTAPRKEEVWDTFKV"
BASE COUNT 629 a 395 c 414 g 419 t
ORIGIN
Query Match 29.5%; Score 1077.8; DB 9; Length 1857;
Best Local Similarity 95.7%; Pred. No. 2.5e-229;
Matches 1119; Conservative 0; Mismatches 27; Indels 3; Gaps 1;
QY 2480 AAACGGAAGAGTCCAGTAGCCTACCCACAGAAAGCCAGAGGAGGAGACAGAAAATT 2539
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 1 AAACGGAAGAGTCCAGTAGCCTACCCACAGAAAGCCAGAGGAGGAGTGCACAGAAAATT 60
QY 2540 CCACACTGGAATCCCTGCTGGGTAAATAAATACTAACACAAAACTAAAGAAGACACGGAGG 2599
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 61 CCACACTGGAATCCCTGCTGGGTAAATAAATACTAACACAAAACTAAAGAAGACACGGAGG 120
QY 2600 CTGAGTCCACGAGTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2659
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 121 CTGAGTCCACGAGTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
QY 2660 CAGAGAAGAAACTGGGCACCCACCAAGCATCGATGTTACAGAAGTCTCTCAGTGTGTCATAG 2719
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 181 CAGAGAAGAGCTTGACACCCACCAAGCTCAATGTTACAGAAGTCTCTCAGTGTGTCATAG 240
QY 2720 CAAGGCCAAGGAGAACTCTTGGATTAGTGGGAAACCCCAACACAGCAGGTGTGGAGG 2779
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 241 CAAGTGCCAAAGGAGAACTCTTGGATTAGTGGGAAACCCCAACACAGCAGGTGTGGAGG 300
QY 2780 AACGCACATAAATCCAGAAACCTTTGCCAAAGATTAAGAGACAAACAGAAATCACTCAA 2839
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 301 AAAGTGCTAAATCCCAAGAAACCTCTGCCAAAGATTAAGAGACAAACAGAAATCACTCAA 360
QY 2840 ATTCTGATAACACAGAGACTAAAGATCTGCCCCCAAACTCAAACTCTGCGGAGGAGC 2899
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 361 ATTCTGATAACACAGAGACTAAAGATCTTACCCCAAACTCAAACTCTGCGGAGGAGC 420
QY 2900 CAAGAAAGCTCAGAAATCTGGGATTATGAACAACAAGGTCACCCCACTGCCA 2959
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 421 CAAGAAAGCTCAGAAATCTGGGATTATGAACAACAAGGTCACCCCACTGCCA 480
QY 2960 ATTCTGACTGAACCCAGGACACCCAGATCAAGGACAACTTTGCACATTTGGGAGGTGT 3019
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 481 ATTCTGACTGAACCCAGGACACCCAGATCAAGGACAACTTTGCACATTTGGGAGGTGT 540
QY 3020 GTCTTTGGGAGTTTATGACCTGACCCCTGGTCTGTGCTTTCAGAAATCAAAAGTTCAA 3079
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 541 GTCTTTGGGAGTTTATGACCTGACCCCTGGTCTGTGCTTTCAGAAATCAAAAGTTCAA 600
QY 3080 AGCAGTATCTATTGTGGCTTCTGAAATGGAGAAAACCCCACTTTTTCCTTAAAGGAGA 3139
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 601 AGCATGTATCTATTGTGGCTTCTGAAATGGAGAAAACCCCACTTTTTCCTTAAAGGAGA 660
QY 3140 AATCTCACCAAGCCTTAAGCAGCTGAGTTTGTGCACAAATCCAATCAGAGCCCATAG 3199
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 661 AATCTCACCAAGCCTTAAGCAGCTGAGTTTGTGCACAAATCCAATCAGAGTGCACAG 720
QY 3200 ATAAGGCTGAAGTATGCTTTGGGAGACCAAGGCGAGTCCATTTTGGAGATGAGAAGC 3259
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 721 ATAAGGCTGAAGTATGCTTTGGGAGATCCAAGGCCAGTCCATTTTGGAGATGAGAAGC 780
QY 3260 TTTTGATTTTCAAGACTCCAGTTTCCAGAGAGGGGCAAGAGAGGAGGAGGTGACG 3319
||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 781 ATTTTCATTTCGAAGACTCCAGTTCTCCGAAGAGGGGCAAGAGAGGAGGAGGTGACG 840
```



```
QY 3320 CTCGTGACGACCAATGTGTGTGGCAGAGCGAAGAACTGCCCCCAAGCTGTAGCAT 3379
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 841 CTCATGCAGCAAAGTCTGTGTGGCAGAGTGAAGAACTGCCCCCAAGCTGTAGCAT 900
QY 3380 CAAAACACAGAGATGAATCTCAACCAATAGGACACAGAGAAAAGACATCTTCTT 3439
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 901 CAAAACACAGAGATGAATCTCAACCAATAGGACACAGAGAAAAGACATCTTCTT 957
QY 3440 CTCGAGAGAAATGTCGTGGCTCTTATAACTCAAGTAATACTCCAGCAACCTTTAAACAT 3499
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 958 TTGAGAGAAATGTCGTGGCTCTTATAACTCAAGTAATACTCCAGCAACCTTTAACT 1017
QY 3500 CAGAGCAGAGGTTTCTCTTGGAGTTTGGAGCCCTCAGCAACCTTCCAGCAACCTTTAACT 3559
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1018 CAGAGCAGAGGTTTCTCTTGGAGTTTGGAGCCCTCAGCAACCTTCCAGCAACCTTTAACT 1077
QY 3560 GTCTAGCTTTACTGCTCTCTGCTCTCAAGTCAATAGTCAATAGTCAAGGCTTCCAGCAACCTTTAACT 3619
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1078 GTCTAGCTTTACTGCTCTCTGCTCTCAAGTCAATAGTCAATAGTCAAGGCTTCCAGCAACCTTTAACT 1137
3620 AGAGATCTGGGATAGTTTAAAGTGTAG 3648
1138 AGAGATCTGGGATAGTTTAAAGTGTAG 1166

RESULT 11
AL772387
LOCUS Mus musculus chromosome 2 clone RP23-133B16, linear HTG 09-JUL-2002
DEFINITION HTGS_ACTIVEFIN; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
ACCESSION AL772387
VERSION 5 GI:21727550
KEYWORDS HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 209802)
Almeida,J.
Direct Submission
Submitted (08-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
On Jul 10, 2002 this sequence version replaced gi:21667969.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: bml33816
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 208844 bases at least Q40
Consensus quality: 208951 bases at least Q30
Consensus quality: 209172 bases at least Q20
Insert size: 209402; sum-of-contigs
Insert size: 194408; 1.3% error; agarose-fp
Quality coverage: 8.01x in Q20 bases; sum-of-contigs Quality
coverage: 8.63x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 11097: contig of 11097 bp in length
* 11098 11197: gap of 100 bp
* 11198 15335: contig of 4138 bp in length
```

```
* 15336 15435: gap of 100 bp
* 15436 78237: contig of 62802 bp in length
* 78238 78337: gap of 100 bp
* 78338 80885: contig of 2548 bp in length
* 80886 80985: gap of 100 bp
* 80986 209802: contig of 128817 bp in length.
FEATURES
Location/Qualifiers
1..209802
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-133B16"
/clone_lib="RPCI-23"
1..11097
/note="assembly_fragment:00793
fragment_chain:1
clone_end:17
vector_side:left"
11198..15335
/note="assembly_fragment:00044
fragment_chain:1"
15436..78237
/note="assembly_fragment:00911
fragment_chain:1"
78338..80885
/note="assembly_fragment:00030"
80986..209802
/note="assembly_fragment:00556"
BASE COUNT 65094 a 40593 c 41610 g 62105 t 400 others
ORIGIN
Query Match 28.1%; Score 1026.6; DB 2; Length 209802;
Best Local Similarity 81.3%; Pred No. 7.7e-218;
Matches 1223; Conservative 0; Mismatches 264; Indels 18; Gaps 2;
QY 2143 CGGAGACGCTGAAATAAATCTATGCCCAACTGGAAATATATAAAGAAGAAGATGATC 2202
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188329 CAGGATGAATGAAATAAATCTATGCTAGTAAATATATAAAGATGATC 188388
QY 2203 ACATAACACCCCACTCCAGAAAAGCGTCTCGAAGAAGGCGCTGCTTCCATC 2262
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188389 ACATAACACCCCACTCCAGAAAAGCGTCTCGAAGAAGGCTTCCATC 188448
QY 2263 ATGAGACGATTTACGAGATCCAGAGACAGTCAGCGGCGCTGCTTAAAGAGGACAG 2322
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188449 ATGAGCGCATCTAGTATCCAGAGACAGTCAGGACAGCTGCTTAAAGAGGACAG 188508
QY 2323 GAGGGCGCGCATGCGCACAGCAAGCACTGCCCTCATCAGGAAGAACCCCGCAGAG 2382
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188509 GAAGGCACAGACAGTCAGCAAGGCGGCGCTGCTCGGAAGAAGCACTACTGAG 188568
QY 2383 TCTTCAGGGAACACAGGAAATCCAGAGGAGAGCCCTGAAAACCGAGTCTTCACTC 2442
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188569 TCTTCAGGGAATACAGGAGAGCCCAAGGAAGAGTCCCTTAAATCGGCTTCTCTC 188628
QY 2443 AAGAAATCCACAGCAGCTTATGACCACTGAGAGACCAACGAGAGTCCAGTACCTTA 2502
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188629 AAGAAATCCACAGCAGCATATGACCATGTGAGAGACCAACTGTAGAAATCCAGAGCTTG 188688
QY 2503 CCCACAGAAAGCAGAGAGGAGACACAGAAAATTCACACTGGAATCCCTGTGGGT 2562
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188689 CCCATGGAGAGCAGAGGAGGAGGCGCCACCGAAATTCACCTTGGAGTCGCTCAAGT 188748
QY 2563 AAAAACTAACACAAAAAATAAAGAAGACAGGAGGCTGAGTCCACGAGTCGGTCCG 2622
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188749 AAAAACTAACACAAAAAAGCTCAAGAAGATAGTGAGCGCGAGTCCACTGAGTCAGTCCG 188808
QY 2623 TTGGTGTCAAGTCAAGAGCGCTCACACCTCAGCTCAGAGAAGAAAATGGGACACCA 2682
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188809 TTGGTGTCAAGTCAAGAGCTGCTCACACCTCAGCTCAGAGAAGAGAGCTGGGACACCA 188868
QY 2683 CGAATCATGCTTACAGAAAGTCTCTAGTGTATAGCAAGGCCCAAGGAGAGACTCTT 2742
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db	188869	CGTACATCTATGTTACAGAAATCTCTCAGTGTCTATTGCAAGTGCCCAAGGAGAGACACTT	188928
QY	2743	GGATTAGCTGGGAAACCCAAACAGCAGCGTGTGGAAGACGCACTAAATCCCGAGAACCT	2802
Db	188929	GGCTGGCTGGCAAAACCCAGACTTTAGTTATGGAAGACCGAGCTAAGTCCCGAGAAA---	188985
QY	2803	TTGCCAAAAGATAAAGAGACAACAGAAATCACTCAAAATTTCTGATAACACAGAGACTAAA	2862
Db	188986	---CCAAAGATAGAGAGACCACTCAGGAAGTACTCAAAATTCAGATAATGTAGAGACTA--	189040
QY	2863	GATCCTGCCCCCAAAACTCAAAATCTCGGAGGAGGCCAAGAAAGCCTCAGAAATCTGGG	2922
Db	189041	-----TCCAAACTCCGGTACATGAGAGGCCAAGAAAGCCCGAGAAATCTGGG	189090
QY	2923	ATTATGAACAACAAGAGGTCAACCCCACTGCAATTTCTGACCTGAACCCAGGCACC	2982
Db	189091	ATTATGAACAACAGAGAGGTGACCTTCCCACTGCCAATCCTGACGTGAGCTCCGGTATC	189150
QY	2983	ACCAGATGAAGGACAACTTTGACATTTGGGAGGTGTCTTGGGAGGTTTATGACCTG	3042
Db	189151	ACCAGATAAAGGACAATTTGATATCGGAAGTATGCCCTTGGGAGGTTTATGACCTG	189210
QY	3043	ACCCCTGCTCTGTGCTTCAGAAATCAAAAGTTCAAAACGACGTATCTATTGTGGCTTCT	3102
Db	189211	ACCCCTGGTCTCTATGCTTCAGAGGCCAAAGCTCAAAAACAGGTATCAATCGCAGCTTCT	189270
QY	3103	GAAATGGAAGAAAACCCCACTTTTCCCTTAAAGGAGAAATCTCACCAAGCCTAAGGCA	3162
Db	189271	GAAATGGAACAACCCCTGCTGCTTTTGAAGGAGAAGTCTTACCACAAAGTCTAAGGCA	189330
QY	3163	GCTGAGGTTTGTGAGCAATCCAATCAGAAAGCGCATAGATAAGGCTGAAGTATGCCCTTGG	3222
Db	189331	ACTGAGGTTCTCTACCAAGCCAATCACAAGACATAGACAACAGAGGTGCCCCCTGG	189390
QY	3223	GAGAGCAAGGCCAGTCCATTTTGGAGATGAGAGCTTTTGAATTTCCAAAGTCCAGTT	3282
Db	189391	GAAATCCAGTCAGTCCCTTTTGAAGATGAGAATCGTCTGATTTCTAAGACTCCAGTT	189450
QY	3283	CTCCAGAGAGGGCAAAAGAGGAGACGAGGTGACGCTCGTCAGCCCAATGTGTGCT	3342
Db	189451	CTCCAGGAGAGCCAGAGAGGAGATGGAAGCCAGCTCTACACACCAATATGTGCT	189510
QY	3343	GGCAGAGCGAAGAACTGCCCCCAAGCTGTAGCATCAAAAACAGAGAAATGAAATCTC	3402
Db	189511	GGCAATATGAAGAACTGCCCCCAAGCTGTAGCACCACCAAAAGTAGAAGATGAAATCTC	189570
QY	3403	AACCAATAGGACACCAGCAAAAAGAGACATCTTCTCTGAGGAGAAATGTGCTGGCTCC	3462
Db	189571	AACCAATGGGAGACCCAGGAGAAACAGACATCTTCTCTGTGGATATCTTCTGGCTCC	189630
QY	3463	TATACTCAAGTAATAACTTCCAGCAACCTTTAAACATCACGAGCAGAGGTTTGTCTTGG	3522
Db	189631	TGTAACCTCAAGTAAATAACTTCCATCAGCCCTTAACATCACGAGCGAAGTGTCTTGG	189690
QY	3523	GAGTTTGAGCCCCAGCTCAACCAATGCTGGAAGAAGTGTAGCTTTACCTGCCCTTCT	3582
Db	189691	GAGTTTGAGCCCCCTAGAACCAACCAATGCTGGAAGAAGGTAACCTTACCTGCCCTCT	189750
QY	3583	GCTCTAAGTGAATAAGATACAGGGCTAGCAAGAAGAGATCTGGGATAGTTTAAA	3642
Db	189751	GCTTTAAGTGCAATAAGATACAGGGCTCAGAAATAGGAAGTGTGGACACTTTTCAA	189810
QY	3643	GTGTA 3647	
Db	189811	GTGGA 189815	
RESULT	12		
LOCUS	AL358932	111117 bp	DNA linear HTG 16-AUG-2001
DEFINITION	Homo sapiens chromosome 10 clone RP11-395P8, *** SEQUENCING IN		
PROGRESS	***, 25 unordered pieces.		
ACCESSION	AL358932		

VERSION	AL358932.3	GI:9930957
KEYWORDS	HTG; HTGS_PHASE1; HTGS_CANCELLED.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	1 (bases 1 to 111117)	
AUTHORS	Direct Submission	
TITLE	Submitted (15-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,	
JOURNAL	CB10 15A, UK. E-mail enquiries: humquery@sanger.ac.uk	
COMMENT	Requests: clonerrequest@sanger.ac.uk	
	On Aug 27, 2000 this sequence version replaced gi:9926667.	
	----- Genome Center	
	Center: Sanger Centre	
	Center code: SC	
	Web site: http://www.sanger.ac.uk	
	Contact: humquery@sanger.ac.uk	
	----- Project Information	
	Center project name: BA395P8	
	----- Summary Statistics	
	Assembly program: XGAP4; version 4.5	
	Sequencing vector: plasmid; L08752; 100% of reads	
	Chemistry: Dye-terminator Big Dye; 100% of reads	
	Consensus quality: 90261 bases at least Q40	
	Consensus quality: 99226 bases at least Q30	
	Consensus quality: 104490 bases at least Q20	
	Insert size: 108717; sum-of-contigs	
	Insert size: 178772; 24.0% error; agarose-1p	
	Quality coverage: 1.84x in Q20 bases; sum-of-contigs Quality	
	coverage: 1.42x in Q20 bases; agarose-1p	
	-----	
	* NOTE: This is a 'working draft' sequence. It currently	
	* consists of 25 contigs. The true order of the pieces	
	* is not known and their order in this sequence record is	
	* arbitrary. Gaps between the contigs are represented as	
	* runs of N, but the exact sizes of the gaps are unknown.	
	* This record will be updated with the finished sequence	
	* as soon as it is available and the accession number will	
	* be preserved.	
	* 1 2287: contig of 2287 bp in length	
	* 2288 2387: gap of 100 bp	
	* 2388 6317: contig of 3930 bp in length	
	* 6318 6417: gap of 100 bp	
	* 6418 9725: contig of 3308 bp in length	
	* 9726 9825: gap of 100 bp	
	* 9826 16607: contig of 6782 bp in length	
	* 16608 16707: gap of 100 bp	
	* 16708 20159: contig of 3452 bp in length	
	* 20160 20259: gap of 100 bp	
	* 20260 23641: contig of 3382 bp in length	
	* 23642 23741: gap of 100 bp	
	* 23742 27399: contig of 3658 bp in length	
	* 27400 27499: gap of 100 bp	
	* 27500 32754: contig of 5255 bp in length	
	* 32755 32854: gap of 100 bp	
	* 32855 35240: contig of 2386 bp in length	
	* 35241 35340: gap of 100 bp	
	* 35341 40948: contig of 5608 bp in length	
	* 40949 41048: gap of 100 bp	
	* 41049 43758: contig of 2710 bp in length	
	* 43759 43858: gap of 100 bp	
	* 43859 48684: contig of 4826 bp in length	
	* 48685 48784: gap of 100 bp	
	* 48785 52405: contig of 3621 bp in length	
	* 52406 52505: gap of 100 bp	
	* 52506 56282: contig of 3777 bp in length	
	* 56283 56382: gap of 100 bp	
	* 56383 59669: contig of 3287 bp in length	
	* 59670 59769: gap of 100 bp	
	* 59770 63200: contig of 3431 bp in length	
	* 63201 63300: gap of 100 bp	
	* 63301 68131: contig of 4831 bp in length	





```

RESULT 14
AC094823/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-5J17, *** SEQUENCING IN PROGRESS ***,
61 unorderded pieces.
AC094823
HTG: HTGS-PHASE1.
KEYWORDS
SOURCE
ORGANISM
Rattus norvegicus.

REFERENCE
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraturge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Blinage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burckett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Morgan,G., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,G., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaiker,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 174215)
Worley K.C.

Direct Submission
Unpublished
2 (bases 1 to 174215)

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15624659.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GBLR
Center clone name: CH230-5J17
----- Summary Statistics

```

```

Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 152461 bases at least Q40
Consensus quality: 158423 bases at least Q30
Consensus quality: 163424 bases at least Q20
Estimated insert size: 155278; sum-of-contigs estimation
Quality coverage: 2.6x in Q20 bases; agarose-fp estimation
Quality coverage: 2.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 61 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
  be preserved.
*
* 1
  7925: contig of 7925 bp in length
  8025: gap of unknown length
  13848: contig of 5823 bp in length
  13849: gap of unknown length
  18634: contig of 4686 bp in length
  18635: gap of unknown length
  25312: contig of 6578 bp in length
  25313: gap of unknown length
  29445: contig of 4533 bp in length
  30445: gap of unknown length
  33954: contig of 3909 bp in length
  34054: gap of unknown length
  40909: contig of 6855 bp in length
  41009: gap of unknown length
  45315: contig of 4306 bp in length
  45415: gap of unknown length
  50962: contig of 5547 bp in length
  50963: gap of unknown length
  51063: contig of 5425 bp in length
  56487: gap of unknown length
  56588: gap of unknown length
  59752: contig of 3165 bp in length
  59852: gap of unknown length
  63689: contig of 3837 bp in length
  63789: gap of unknown length
  67805: contig of 4016 bp in length
  67905: gap of unknown length
  71547: contig of 3642 bp in length
  71548: gap of unknown length
  71648: contig of 4636 bp in length
  76284: gap of unknown length
  76384: contig of 3566 bp in length
  79950: gap of unknown length
  80049: gap of unknown length
  84087: contig of 4038 bp in length
  84088: gap of unknown length
  87030: contig of 2843 bp in length
  87130: gap of unknown length
  90723: contig of 3593 bp in length
  90823: gap of unknown length
  93351: contig of 2528 bp in length
  93352: gap of unknown length
  97019: contig of 3568 bp in length
  97119: gap of unknown length
  100214: contig of 3095 bp in length
  100215: gap of unknown length
  103569: contig of 3255 bp in length
  103570: gap of unknown length
  105874: contig of 2205 bp in length
  105875: gap of unknown length
  108068: contig of 2094 bp in length
  108168: gap of unknown length
  110705: contig of 2537 bp in length
  110706: gap of unknown length
  113382: contig of 2577 bp in length
  113482: gap of unknown length

```

```
* 113483 116314: contig of 2832 bp in length
* 116315 116414: gap of unknown length
* 116415 118362: contig of 1948 bp in length
* 118363 118462: gap of unknown length
* 118463 121415: contig of 2933 bp in length
* 121416 121515: gap of unknown length
* 121516 123174: contig of 1659 bp in length
* 123175 123274: gap of unknown length
* 123275 124680: contig of 1406 bp in length
* 124681 124780: gap of unknown length
* 124781 126634: contig of 1854 bp in length
* 126635 126734: gap of unknown length
* 126735 128743: contig of 2009 bp in length
* 128744 128843: gap of unknown length
* 128844 130748: contig of 1905 bp in length
* 130749 130848: gap of unknown length
* 130849 132599: contig of 1751 bp in length
* 132600 132699: gap of unknown length
* 132700 134625: contig of 1926 bp in length
* 134626 134725: gap of unknown length
* 134726 137010: contig of 2285 bp in length
* 137011 137110: gap of unknown length
* 137111 138401: contig of 1291 bp in length
* 138402 138501: gap of unknown length
* 138502 140456: contig of 1955 bp in length
* 140457 140556: gap of unknown length
* 140557 142606: contig of 2050 bp in length
* 142607 142706: gap of unknown length
* 142707 144024: contig of 1318 bp in length
* 144025 144124: gap of unknown length
* 144125 145788: contig of 1664 bp in length
* 145789 145888: gap of unknown length
* 145889 147892: contig of 2004 bp in length
* 147893 149008: contig of 1016 bp in length
* 149009 149108: gap of unknown length
* 149109 150676: contig of 1568 bp in length
* 150677 150776: gap of unknown length
* 150777 152366: contig of 1590 bp in length
* 152367 152466: gap of unknown length
* 152467 154435: contig of 1969 bp in length
* 154436 154535: gap of unknown length
* 154536 155809: contig of 1274 bp in length
* 155810 155909: gap of unknown length
* 155910 157454: contig of 1545 bp in length
* 157455 157554: gap of unknown length
* 157555 159422: contig of 1868 bp in length
* 159423 159522: gap of unknown length
* 159523 161128: contig of 1606 bp in length
* 161129 161228: gap of unknown length
* 161229 162609: contig of 1381 bp in length
* 162610 162709: gap of unknown length
* 162710 164067: contig of 1358 bp in length
* 164068 164167: gap of unknown length
* 164168 165531: contig of 1364 bp in length
* 165532 165631: gap of unknown length
* 165632 167163: contig of 1532 bp in length
* 167164 167263: gap of unknown length

Query Match      20.2%; Score 736.6; DB 2; Length 174215;
Best Local Similarity 88.5%; Pred. No. 3.3e-153;
Matches 799; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 1 ATGGGAGCATGGCTTACCCCTTACTCTCTCCCTCCCTCCCTGCTTGCCTCAGCTGGGATGGGA 60
Db 17025 ATGGGAGCATGGCTTACTCTTTTACTCTCTCCCTCCCTCCCTGCTTGCCTCAGCTGGGATGGGA 16966
QY 61 GCTGTTGGCCGAGCCGCGGACCCCAAGACGCGCGGATTCCCTCCGAGAGAGACCCCG 120
Db 16965 GAGTTTGGTCCAGCTCGACCTTCAGACGCGCCAGATCTTCTCGAGAGAGACTTCG 16906
QY 121 AAGGGAGACCCGACGCCCGGAGCTCGAGCTCTGCTCGAGCTCTCTCGGCTCC 180
```

```
Db 16905 AGGGGAAGCAGCAGCGTTCAGAGCTGCGCGAGGCTCTGCCCGGACCCCTCTATTCC 16846
QY 181 TGGAGCCGCTCCACGATGGCGACCATCTTGGCGCAGAACTCGCGAGAGCTGCCCATG 240
Db 16845 TGGAGCCGCTCCACGATGGCGACCATCTTGGCACAAAATAGCCGAGGAGTGCCCATG 16786
QY 241 GACGTGGGCTCTTACTCTTACACCGGGGACTCCACCAGCTTGAAGCGACCCAACTGCTCC 300
Db 16785 GACGTGGGCTCTTACTCTTATACTGGGACTTCCACCAGCTGAAGAGACCAACTGCTCA 16726
QY 301 GCGCGCTAGACTTGGCGGCTGCGCGGAAAGTGGCCAGCCCTGGCCAGCGCACCC 360
Db 16725 GCGCGCTAGACTTGGCGGCTGCGCGGAAAGTCCAGCTCCCTAGCCAGCTCCCATCT 16666
QY 361 TCCTTGCACCGGGGCTGGACACTGACACAGCCACCAACTTCTCAACTGATGCTG 420
Db 16665 TCCTTGCACGGGGCTGGACACTCTGACACAGCCACCAACTTCTCAACTGATGTTG 16606
QY 421 CAGAGCAATAAGTTCGGGAGCAGAACTTGCAGGACGACTTGATTTGGTACCAGGCGCTG 480
Db 16605 CAGAGCAATAAGTTCGGGAGCAGACCTGTCAGAGTACCTGCTGATGGTACCAGGCTTG 16546
QY 481 GTGTGAGCTTCTGGAGGGGAGCCAGCATCTCCCGGGGCGCCATCACCCTTACGAC 540
Db 16545 GTGCGAAGTCTCTGGAGGGAGAGCCAGCATCTCCCGGGGCGCCATCACCCTTACGAC 16486
QY 541 GATTGCTGTCGCGACCGCCACAGGTCTTCTCCAGGCCACCGCGGAGAGAGCCG 500
Db 16485 GATTGCTGTCGCGACCGCCCGCCGCAAGTCTTCTCCAGGCCACCGCGTGGAGAGTGGC 16426
QY 601 ATCTGCTCCAAAGACTGTCTCTCCCGCACCCACCTTGGCCAGCCACTCTGGAGACC 660
Db 16425 ATCTGCTGTCAGGACTGTCTCTCCCGCTCAGCTGCTTACGCCACTCTGGAGACC 16366
QY 661 GAGTGGTCCACGGGCTCCGGCGCAAGTGGAGGCGCCACTTACACCGCGCGGCCCAAT 720
Db 16365 GAGTGGTCCACGGGCTTCGCGCAAGTGGAGGCGCCACTTACACCGCGCGGCCCAAT 16306
QY 721 CAGGGCGCGGGGCTGGGCCACAGCTGGCGGCACAGGAGGCTGGCGGGGACAG 780
Db 16305 CAGGGCGCGGGGCTGGGCCACAGCTGGCGGCACAGGAGGCTGGCGGGGACAG 16246
QY 781 AGCCACTTCAAGTGTCTCCGCTTATCTGGAGTGCAGAGAGCGGAGTTACAAAGCCCGG 840
Db 16245 AGCCATGTCAAGTGTCTCCACCTTTCTGGAGTGCAGAGATGGGAGTTACAAAGCCCGG 16186
QY 841 TGGCTGGTACTCTTCTCTGCTGCTTACGAGTGGAGGCTTACCTGCTGCGGAAATTC 900
Db 16185 TGGCTGGTACTCTCTCTGCTGCTTCTACGGGCTTCAACCTTACCTGCTGCGGAAATTC 16126
QY 901 AGG 903
Db 16125 AGG 16123

RESULT 15
LOCUS AC109216/3
DEFINITION Mus musculus clone RP23-390E1, WORKING DRAFT SEQUENCE, 13 ordered
pieces.
ACCESSION AC109216
VERSION AC109216.3 GI:20336087
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 163199)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-390E1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 163199)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
```

Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collimore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# Direct Submission

Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 163199)

## TITLE JOURNAL REFERENCE AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collimore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# Direct Submission

Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Apr 28, 2002 this sequence version replaced gi:20303799.

All repeats were identified using RepeatMasker:

Smit, A. F. A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L20681

Center clone name: 390\_E\_1

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 159494 bases at least Q40

Consensus quality: 160888 bases at least Q30

Consensus quality: 161312 bases at least Q20

Insert size: 168000; agarose-fp

Insert size: 161999; sum-of-contigs

Quality coverage: 8.6 in Q20 bases; agarose-fp

Quality coverage: 8.9 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor.  
\* This sequence will be preserved  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 3749: contig of 3749 bp in length  
\* 3750 3849: gap of 100 bp  
\* 3850 4940: contig of 1091 bp in length  
\* 4941 5040: gap of 100 bp  
\* 5041 6506: contig of 1466 bp in length  
\* 6507 6606: gap of 100 bp  
\* 6607 7378: contig of 772 bp in length  
\* 7379 7478: gap of 100 bp  
\* 7479 9982: contig of 2504 bp in length  
\* 9983 10082: gap of 100 bp  
\* 10083 11915: contig of 1833 bp in length  
\* 11916 12015: gap of 100 bp  
\* 12016 14172: contig of 2157 bp in length  
\* 14173 14272: gap of 100 bp  
\* 14273 18737: contig of 4465 bp in length  
\* 18738 18837: gap of 100 bp  
\* 18838 88219: contig of 69382 bp in length  
\* 88220 88319: gap of 100 bp  
\* 88320 101683: contig of 13364 bp in length  
\* 101684 101783: gap of 100 bp  
\* 101784 123433: contig of 21650 bp in length  
\* 123434 123533: gap of 100 bp  
\* 123534 154328: contig of 30795 bp in length  
\* 154329 154428: gap of 100 bp  
\* 154429 163199: contig of 8771 bp in length.

## FEATURES Source

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP23-390E1"  
/clone\_lib="RPCI-23 Female Mouse BAC"

## misc\_feature

/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"

## misc\_feature

3850..4940  
/note="assembly\_fragment"

## misc\_feature

5041..6506  
/note="assembly\_fragment"

## misc\_feature

6607..7378  
/note="assembly\_fragment"

## misc\_feature

7479..9982  
/note="assembly\_fragment"

## misc\_feature

10083..11915  
/note="assembly\_fragment"

## misc\_feature

12016..14172  
/note="assembly\_fragment"

## misc\_feature

14273..18737  
/note="assembly\_fragment"

## misc\_feature

18838..88219  
/note="assembly\_fragment"

## misc\_feature

88320..101683  
/note="assembly\_fragment"

## misc\_feature

101784..123433  
/note="assembly\_fragment"

## misc\_feature

123534..154328  
/note="assembly\_fragment"

## misc\_feature

154429..163199  
/note="assembly\_fragment"

clone\_end:77  
vector\_side:right"

BASE COUNT 48783 a 31895 c 31334 g 49958 t 1229 others  
ORIGIN





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2003, 04:19:08 ; Search time 695 Seconds  
(without alignments)  
11820.561 Million cell updates/sec

Title: US-09-775-181-1

Perfect score: 3648

Sequence: 1 atgggagccatggcttacc.....gggatagttttaagttag 3648

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*

1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3648	100.0	3648	22	AAD12292 Human G-protein co
2	3525	96.6	7148	22	AAD12294 Human G-protein co
3	1756.4	48.1	1800	22	AAD12293 Human G-protein co
4	342.2	9.4	368	21	AAA42817 Human secreted exp
5	240	6.6	398	23	ABV17708 Human prostate exp
6	240	6.6	470	23	ABV47501 Human prostate exp
7	156.8	4.3	182	21	AAA41583 Human secreted exp
8	64	1.8	83	21	AAC08885 Human secreted pro
9	62.8	1.7	795	23	ABL20735 Drosophila melanog

10	59.8	1.6	2010	22	AAS57193	CDNA encoding Dros
11	59.8	1.6	2010	23	ABL28083	Drosophila melanog
12	56.4	1.5	10732	21	AAA10594	Gene encoding a su
13	52	1.4	1035	21	AAA51616	HIV gp41 coding re
14	52	1.4	2466	21	AAA51614	HIV gp160 coding r
15	52	1.4	2547	21	AAA51615	HIV gp160 with sig
16	52	1.4	2550	24	ABL39955	Synthetic Env poly
17	51.8	1.4	1587	22	AAI61054	Human polynucleoti
18	51.8	1.4	1638	22	AAI59268	Human polynucleoti
19	51.8	1.4	1669	22	AAH17479	Human cDNA sequenc
20	51.2	1.4	1459	21	AAH69799	Human breast tumou
21	51.2	1.4	2288	22	AAI58148	Human polynucleoti
22	51.2	1.4	2549	22	AAS26890	Human cDNA encodin
23	51.2	1.4	3867	22	AAF28253	Human TANGO 275 op
24	51.2	1.4	4225	22	AAF28252	Human TANGO 275 DN
25	48.6	1.3	752	22	AAH04810	Human cDNA clone (
26	48	1.3	2571	20	AAZ08742	HIV SYNgp-160mm co
27	48	1.3	2571	21	AAA93974	Human immunodefici
28	48	1.3	2571	22	AAH43701	Codon optimised HI
29	48	1.3	2571	22	AAH43701	Nucleotide sequenc
30	47.8	1.3	6460	24	ABK91482	Modified HIV prote
31	47.8	1.3	6532	24	ABK91517	Modified HIV prote
32	47.8	1.3	6538	24	ABK91478	Modified HIV prote
33	47.8	1.3	6541	24	ABK91479	Modified HIV prote
34	47.8	1.3	6610	24	ABK91515	Modified HIV prote
35	47.8	1.3	6613	24	ABK91516	Modified HIV prote
36	47.8	1.3	6619	24	ABK91472	Modified HIV prote
37	47.8	1.3	6691	24	ABK91507	Modified HIV prote
38	47.8	1.3	7711	24	ABK91497	Modified HIV prote
39	47.8	1.3	7789	24	ABK91493	Modified HIV prote
40	47.8	1.3	7792	24	ABK91494	Modified HIV prote
41	47.8	1.3	7870	24	ABK91487	Modified HIV prote
42	46.8	1.3	5034	23	ABL26319	Modified HIV prote
43	46.4	1.3	2481	17	AAH13557	Drosophila melanog
44	46.4	1.3	2481	18	AAH73952	Synapl60mm. Synth
45	46.4	1.3	2481	19	AAV23291	HIV-1 gp160 Synth
						Synthetic HIV-1 gp

#### ALIGNMENTS

RESULT 1

AAD12292  
ID AAD12292 standard; cDNA; 3648 BP.

XX AC AAD12292;

XX DT 16-OCT-2001 (first entry)

XX DE Human G-protein coupled receptor (NGPCR) coding sequence #1 cDNA.

XX KW Human: G-protein coupled receptor; NGPCR; gene therapy; cancer; diabetes;

XX KW Inflammation; immune disorder; heart disease; obesity; coronary disease;

XX KW Metabolic disorder; physiological disorder; therapeutic; drug screening;

XX KW signal transduction; behavioural disorder; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 1..3648

XX FT /\*tag= a

XX FT /product= "Human G-protein coupled receptor"

XX PN WC200157086-A2.

XX XX 09-AUG-2001.

XX PF 01-FEB-2001; 2001WO-US03648.

XX XX 04-FEB-2000; 2000US-0180414.

XX PR Human secreted exp

XX PA (LEXI-) LEXICON GENETICS INC.

XX XX













QY 1021 AAAGCCTAGGATTCGTTCTTGAGCCTATGATGCAATTTGCAAGCAGGATTCATCAT 1080  
|||||  
Db 1021 AAAGCCTAGGATTCGTTCTTGAGCCTATGATGCAATTTGCAAGCAGGATTCATCAT 1080  
QY 1081 CCTGGAGTCTTACAGTGAACAACTTTCGGAGAAAGGGTCCGGATCAGATATTTCCAGGA 1140  
Db 1081 CCTGGAGTCTTACAGTGAACAACTTTCGGAGAAAGGGTCCGGATCAGATATTTCCAGGA 1140  
QY 1141 AGTACAAAAGATGTGTGACAAAGCCTATGTCGTGCTACCTTGACAGGAGGCTGCCCC 1200  
Db 1141 AGTACAAAAGATGTGTGACAAAGCCTATGTCGTGCTACCTTGACAGGAGGCTGCCCC 1200  
QY 1201 TTCTGTGCTGATCAGACGCCATGCTTCGTCAGGAAGATAAGTATTTAGCACTTGCCATC 1260  
Db 1201 TTCTGTGCTGATCAGACGCCATGCTTCGTCAGGAAGATAAGTATTTAGCACTTGCCATC 1260  
QY 1261 ATCTCCTCCAGGCTGTGTATGCTGCTGACATGCTGTAGCATGCTGTGTGTACAC 1320  
Db 1261 ATCTCCTCCAGGCTGTGTATGCTGCTGACATGCTGTGTAGCATGCTGTGTGTACAC 1320  
QY 1321 TTTCCGAAAGCAAGAGATCCGGGATCGGCGCTATCCTCTTGGAAACGATCCTTTT 1380  
Db 1321 TTTCCGAAAGCAAGAGATCCGGGATCGGCGCTATCCTCTTGGAAACGATCCTTTT 1380  
QY 1381 GGATCTCTGCTCTATACATTTCCAGTGTATTTTGTACTTTGAGCAAGCACATTTCCG 1440  
Db 1381 GGATCTCTGCTCTATACATTTCCAGTGTATTTTGTACTTTGAGCAAGCACATTTCCG 1440  
QY 1441 TGTATTTCTCTAAGATGGGCTCGTCTTCGCTGTTTGTACTTTGAGCAAGCACATTTCC 1500  
Db 1441 TGTATTTCTCTAAGATGGGCTCGTCTTCGCTGTTTGTACTTTGAGCAAGCACATTTCC 1500  
QY 1501 CTCAAACTTTCAGAGGTTTGAAGGTGTTCTTTCACGAACGCTCAACGAATTCATAT 1560  
Db 1501 CTCAAACTTTCAGAGGTTTGAAGGTGTTCTTTCACGAACGCTCAACGAATTCATAT 1560  
QY 1561 ATGACTGGCGGACGGTCTATGAGGATGCTGGCAGTAATCTCTGCTAGTGTGTTGTT 1620  
Db 1561 ATGACTGGCGGACGGTCTATGAGGATGCTGGCAGTAATCTCTGCTAGTGTGTTGTT 1620  
QY 1621 CTCATTTGGCTGGACTTCATCTGTGTGCCAGATTTGGAGAACAGATTTTCATTTGGC 1680  
Db 1621 CTCATTTGGCTGGACTTCATCTGTGTGCCAGATTTGGAGAACAGATTTTCATTTGGC 1680  
QY 1681 CAGGGGAAACATCCGATCACCTTCATCTCAATATGCTGCTCATTGACCGCTGGGACTAC 1740  
Db 1681 CAGGGGAAACATCCGATCACCTTCATCTCAATATGCTGCTCATTGACCGCTGGGACTAC 1740  
QY 1741 ATGACAGCAGTTGGTATGTGGTCACTTGTGTTTATTCCTCTTGTGGGGTGT 1782  
Db 1741 ATGACAGCAGTTGGTATGTGGTCACTTGTGTTTATTCCTCTTGTGGGGTGT 1782

RESULT 4

AAA42817  
ID AAA42817 standard; cDNA; 368 BP.

AC AAA42817;

XX 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:1557.

Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;  
expressed sequence tag; EST; probe; chemotactic; proliferative;  
immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
antiviral; antidiabetic; antiasthmatic; vulnerary; antiparkinsonian;  
antitumor; osteoprotective; neuroprotective; nootropic; antipsoriatic;  
cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;

lupoid cell deficiency; burn; osteoporosis; osteoarthritis;  
central nervous system disorder; Alzheimer's disease; stroke;  
Parkinson's disease; Huntington's disease; coagulation disorder;  
haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
tumour; infection; depression; psoriasis; ss.  
Homo sapiens.  
WO200021990-A1.  
XX 20-APR-2000.  
XX 15-OCT-1999; 99WO-US24205.  
XX 15-OCT-1998; 98US-0104435.  
XX (GEM ) GENETICS INST INC.  
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M;  
XX WPI; 2000-317937/27.  
XX Isolated polynucleotides, and encoded proteins, comprising secreted  
expressed sequence tags (seSTs), useful for treating various disorders  
such as autoimmune, infectious, and central nervous system disorders -  
Claim 1; Page 487; 618pp; English.  
XX AAA1261 to AAA43419 represent specifically claimed secreted expressed  
sequence tags (seSTs), isolated from human, mouse, xenopus and rat  
tissue sources. The seSTs can have a range of activities depending on  
the tissues they were isolated from. The activities include:  
chemotactic; proliferative; immunomodulatory; haematopoietic;  
chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
antiasthmatic; vulnerary; antitumor; osteopathic; neuroprotective;  
nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;  
anticonvulsant; and antidepressant. The seSTs can be used for gene  
therapy and in vaccines. The seSTs are useful as probes for the  
identification and isolation of full-length cDNAs and genomic DNA  
molecules which correspond to the seSTs. Proteins encoded by the seSTs  
are useful in assays for determining biological activity and raising  
antibodies. They may be useful for treatment of autoimmune disorders  
(multiple sclerosis, insulin dependent diabetes), allergic conditions  
(asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
osteoporosis, osteoarthritis, central nervous system disorders  
(Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation  
disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's  
disease), tumours, bacterial, fungal or viral infections, depression and  
psoriasis. AAA43420 to AAA43425 represent linker variants which are given  
in the exemplification of the present invention.  
XX Sequence 368 BP; 115 A; 78 C; 79 G; 96 T; 0 other;  
SQ Query Match 9.4%; Score 342.2; DB 21; Length 368;  
Best Local Similarity 99.1%; Pred. No. 3e-83;  
Matches 344; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1871 CTGCTATATTCACATTCAGATTTGTTCTTGGCTCAAGACTTCAGTCTGATGATG 1930  
Db 19 CTACTATATTCACATTCAGATTTGTTCTTGGCTCAAGACTTCAGTCTGATGATG 78  
QY 1931 TCATGCTGTATTTTCACATTCAGATTTGTTCTTGGCTCAAGACTTCAGTCTGATGATG 1990  
Db 79 TGATGCTGTATTTTCACATTCAGATTTGTTCTTGGCTCAAGACTTCAGTCTGATGATG 138  
QY 1991 TTCCAAAGTTTTCACATTCAGATTTGTTCTTGGCTCAAGACTTCAGTCTGATGATG 2050  
Db 139 TTCCAAAGTTTTCACATTCAGATTTGTTCTTGGCTCAAGACTTCAGTCTGATGATG 198  
QY 2051 AGGATGAGCTAGATGAGGCGGATCTGGATCTTACCTGAACAGCATATCAATTCAGCT 2110  
Db 1991 AGGATGAGCTAGATGAGGCGGATCTGGATCTTACCTGAACAGCATATCAATTCAGCT 2110

Db 199 AGGATGAGCTAGACATGGGCCGATCTGGATCCTACCTGAACAGCAGTCAATTCAGCCT 258  
QY 2111 GGAGTCAGCACAGCTTGGATCCAGAGGACATTGGGACGAGCTGAAAAAATCTATGCC 2170  
Db 259 GGAGTCAGCACAGCTTGGATCCAGAGGACATTGGGACGAGCTGAAAAAATCTATGCC 318  
QY 2171 AACTGGAATATATAAAGAAAGAGATGATCACAACAAACCCAC 2217  
Db 319 AACTGGAATATATAAAGAAAGAGATGATCACAACAAACCCCTC 365

RESULT 5  
ABV17708  
ID ABV17708 standard; cDNA; 398 BP.  
XX  
AC ABV17708;  
XX  
DT 13-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 17699.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW Pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
PS Claim 1; Page 2926-2927; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 398 BP; 122 A; 82 C; 103 G; 91 T; 0 other;

Query Match 6.6%; Score 240; DB 23; Length 398;  
Best Local Similarity 100.0%; Pred. No. 3.3e-55;  
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 GCGGGGCAAGAGCCACTTCAAGTGGTCTCCGCTTATCTGGAGTGCAGAACGGGAGT 828  
Db 22 GCGGGGCAAGAGCCACTTCAAGTGGTCTCCGCTTATCTGGAGTGCAGAACGGGAGT 81  
QY 829 TACAAGCCGGGTGGTGGTTACTCTTTCTCTGCAATCTAGGGTTGAGCCTTACCTG 888  
Db 82 TACAAGCCGGGTGGTGGTTACTCTTTCTCTGCAATCTAGGGTTGAGCCTTACCTG 141  
QY 889 GTCCCGGAATTCAGGGGTGTCATGAAGTTGACATAAATCTTCAGAAAAGTGCACATTGAC 948  
Db 142 GTCCCGGAATTCAGGGGTGTCATGAAGTTGACATAAATCTTCAGAAAAGTGCACATTGAC 201  
QY 949 CAATGCTCAAGTGAATGGTGGTTTTCAGGAACCTCAATAATGCCACTCAACAATTCAGAG 1008  
Db 202 CAATGCTCAAGTGAATGGTGGTTTTCAGGAACCTCAATAATGCCACTCAACAATTCAGAG 261

RESULT 6  
ABV47501  
ID ABV47501 standard; cDNA; 470 BP.  
XX  
AC ABV47501;  
XX  
DT 16-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker cDNA 47492.  
XX  
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
KW Pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 25-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
PS Claim 1; Page 9338-9339; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX





XX Claim 1; SEQ ID 12960; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. No ORF has yet been conclusively

CC identified within the present sequence. The 5' ESTs were prepared from

CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

CC sequences usually correspond mainly to the 3' untranslated region (UTR)

CC of the mRNA because they are often obtained from oligo-dT primed cDNA

CC libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.

CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.

CC They are used to obtain upstream regulatory sequences and to design

CC expression and secretion vectors.

XX

XX Sequence 83 BP; 27 A; 19 C; 26 G; 11 T; 0 other;

XX

XX Query Match 1.8%; Score 64; DB 21; Length 83;

XX Best Local Similarity 100.0%; Pred. No. 2.2e-07;

XX Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX 2263 ATGAGACGCGATTACGAGATCCAGAGACAGTCAGCGCGAGTGCTCTAAAGAGGACAAG 2322

DB 1 ATGAGACGCGATTACGAGATCCAGAGACAGTCAGCGCGAGTGCTCTAAAGAGGACAAG 60

XX

XX 2323 GAGG 2326

DB 61 GAGG 64

XX

XX RESULT 9

XX ABL20735

XX ID ABL20735 standard; DNA; 795 BP.

XX AC ABL20735;

XX

XX 26-MAR-2002 (first entry)

XX

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 13678.

XX

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ds.

XX

XX Drosophila melanogaster.

XX

XX WO200171042-A2.

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09231.

XX

XX 23-MAR-2000; 2000US-191637P.

XX

XX 11-JUL-2000; 2000US-0614150.

XX

XX (PEKE ) PE CORP NY.

XX

XX Venter JC, Adams M, Li PWD, Myers EW;

XX

XX WPI; 2001-656860/75.

XX

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX

XX Claim 1; SEQ ID NO 13678; 21pp + Sequence Listing; English.

XX

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (AB57737-AB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

XX Sequence 795 BP; 146 A; 250 C; 220 G; 179 T; 0 other;

XX

XX Query Match 1.7%; Score 62.8; DB 23; Length 795;

XX Best Local Similarity 50.3%; Pred. No. 1.8e-06;

XX Matches 154; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

XX

XX 1683 GGGGAAAACATCCGATCACCTCATCTTCAATATGTGCTCATTTGACCGCTGGGACTACAT 1742

DB 246 GGGAGCGCGCGGACGCTGAAGCGTTCCTCTCAAGACGAGCTGGTGGGACTACAC 305

XX

XX 1743 GACAGCAGTTGCTGAATTTTATCTTGTGGGTGTTTATCTCTGCTATGACGTGCG 1802

DB 306 GTTCACCTCGATGGAGGTCCTGTCTCTGGCGCTGGGCGTTTCGCTGTCATCATGCTGCG 365

XX

XX 1803 GACAGTCCCATCGGCAATCCATGAGCCCGCTATATGCTGTTGGAGTTTCACAAATGAGCT 1862

DB 366 GAAGCGCGCGTGGAGTTTCAACGAGAGCGCTTCATCTCGATGGCCATCTACAATGAGTT 425

XX

XX 1863 CATCATCTCTGCTATATATCCATACAAATTAGATTTGTTCTTGCTCAAGACTTCAGTCTGA 1922

DB 426 CTGCTCACCTGCTTCCCTCAAGCTTCCATGCTCTTCTGCGAGTCGCCGCCAATCCGGA 485

XX

XX 1923 TTGGATCTTCATGCTGTATTTTGCACATACATCTATTTGACTGTGACAGTACCATTGGGTT 1982

DB 486 TTTCGTGTACATCATCTTTTCTGCGACACCCAGCTACGGTGACCCCTGCTCTCTGCGCCCT 545

XX

XX 1983 GCTTTT 1988

DB 546 CATCTT 551

XX

XX RESULT 10

XX AAS57193

XX ID AAS57193 standard; cDNA; 2010 BP.

XX

XX AAS57193;

XX

XX 16-JAN-2002 (first entry)

XX

XX cDNA encoding Drosophila G-protein coupled receptor, GCPR #61.

XX

XX Drosophila; G-protein coupled receptor; GCPR; insecticide; diagnostic;

XX mutation detection; ss.

XX

XX Drosophila melanogaster.

XX

XX WO200170980-A2.

XX

XX 27-SEP-2001.

XX

XX 23-MAR-2001; 2001WO-US09341.

XX

XX 23-MAR-2000; 2000US-191638P.

XX

XX 18-JUL-2000; 2000US-0618893.

XX

XX (PEKE ) PE CORP NY.

XX

XX Cravchik A;

XX

XX WPI; 2001-616405/71.

XX

XX P-PSDB; AAU38983.

XX

XX Sixty six Drosophila Melanogaster G-protein coupled receptors (GPCR),

XX useful in the treatment and diagnosis of GPCR-related conditions and

XX for identifying GPCR modulators for use as insecticides -

XX PS Claim 4; Page 275; 392pp; English.

CC The invention relates to sixty six novel isolated Drosophila

CC melanogaster G-protein coupled receptors (GPCR). The GPCR proteins and

CC nucleic acids are useful in the treatment and diagnosis of GPCR-related

CC conditions. The GPCR proteins and nucleic acids are also useful for

CC identifying modulators of GPCR proteins for use as insecticides. The

CC nucleic acid can also be used to detect mutations in GPCR genes and gene

CC expression products such as mRNA. AAS57072-AAS57203 represent D.

CC melanogaster G-coupled protein receptor genomic and coding sequences of

CC the invention.

XX SQ Sequence 2010 BP; 410 A; 590 C; 571 G; 439 T; 0 other;

Query Match 1.6%; Score 59.8; DB 22; Length 2010;

Best Local Similarity 43.8%; Pred. No. 2.1e-05;

Matches 445; Conservative 0; Mismatches 542; Indels 30; Gaps 3;

QY 1229 TCAGAGAGATAAGTATTTACGACTTGGCCATCATCTCCCTTCCAAAGCCCTGTGTATGCTGC 1288

779 TCAATGTGCGCGCTGCTCGCCCTCTGGTGGCCATCGTACTGGCGCTGCATCTGT 838

QY 1289 TCGACTTCGTTAGCATGTGGTGGTGTACACTTTGGCAAAGCAAGACATCCGGGCGAT 1348

Db 839 GGTGATTTGCTCTGGGGTGATTTGTTCCGGCAGCGAAAGTGCAAGGCCATTGCGTGTG 898

QY 1349 CGGGCCTTATCTGTGGAAAGACCTTTTGGATCTGCTCTTATCTTCCAGTTG 1408

Db 899 GTATGTGGAGTGTGCTGGAGAGTACTGCTGGGCATTTGTTTACTTTATGTCATCTGTG 958

QY 1409 TTATTTGTACTTTGAGCCAGCACATTTTCGCTGTATTTCTCTAAGATGGGCTGCTTTC 1468

Db 959 CGTCCATTTCTCCCGCATCCAGGAGCGTCTCTTGGAGCCCTGGCTTCGGGAGC 1018

QY 1469 TCGGTTTTCGCTACTGTTTACGAACTGTCTACTCTCAAACTTCACAGGGTTTTGAAGGTG 1528

Db 1019 TGGGCTTTCATCACTGCTACGGCGCCATCATCTAAGTTGTACCCGCCACCTGGTGACT 1078

QY 1529 TTCTTTTCAGAACGGCTCAACGAATTCATATATGACTGGCGAGCGGTTCATGAGATGC 1588

Db 1079 TTGCGACCCGGAAGCGCATCG---CTGGGTGCTGGCGAGCTGGACCTGCTCAAGTATC 1135

QY 1589 TGGCAGTAACTACTTGGTAGTGTGTTTGGTTCTCATTTGGCTGGAGTTCATCTGTGTGCC 1648

Db 1136 TGGGCACCATGCTTTGGCGTCACTGTCTATAGCCGCCCTTACGGCTCGTCTGTGG 1195

QY 1649 AGAATTTGGAGAAACAGATTTCACTTATTTGGCCAGGGGAAAACATCCGATCACCTCATCT 1708

Db 1196 ATCTTCTCGA-----AAGTGGCGAGCTGGAGAGCCCTCAGGGAGCGGAC 1240

QY 1709 TCAATATGTGCTCATGTAGCCGCTGGGACTACATGACACAGTGTGCTGAATTTTATTC 1768

Db 1241 CGAACACCTTGCCATCCGCTCAAGTGGAACTGGTCAACGAGACACCGAGATGCTCATCC 1300

QY 1769 TCTTGTGGGGTGTATATCTCTGCTATGCACTGGGACAGTCCCATCGGCAATTCATGAGC 1828

Db 1301 TGTGCTTTGGAGTGCACCTGTCCATCGCCAGTCGGAAATGCCAACACCCAGTTCGGGAGC 1360

QY 1829 CCGCTATATGCGTGTGGAGTTCACAATGAGCTCATCATCTCTGCTATATTTCCATACAA 1888

Db 1361 GACAAATTTCTGTCACCGCACTGACGCTGGAGTTCCTGGTCTCGTCCAGCTTCTACTTTC 1420

QY 1889 TTAGATTTGCTTCTGCTCAAGACTTCATCTGATTTGGATGTTGATGTTTATTTGAC 1948

Db 1421 TGGCTTTTCTACTACCTGCGCGAAATGAGTCCCGACGCCCATCTGTGCTGGCCCTGTTTCATCC 1480

QY 1949 ATACTCATTTGACTGTGACAGTCAACATTTGGTGTGCTTTTGTATTTCCAAAGTTTTCACATT 2008

Db 1481 GCTCCCGAGCTGAGGAATAGTTTTCGGCTTGGGCTGTGATATTTGGCCAAAGTTGT----- 1534

QY 2009 CAAGCAATACCCACGAGATGATTTGCTTACAGAAGCATATGAGGATGAGCTAGACATGG 2068

Db 1535 -----GGTACCAGCACAAAGCAGGGTACCTCACAGCACCGCCAGCGGCTGGCGGAG 1588

QY 2069 GCCGATCTCGATCCTACCTGAACAGCAGCATATCAATTTACGCTGGAGTGAGCACAGCTTGG 2128

Db 1589 GATATGCCGGGCTGTCTTGGGGATCCGGACATTTGGTGAGCTGACCATATCCGAATGA 1648

QY 2129 ATCCAGAGACATTTGGGACGAGCTGAAAGAACTCTATCCCAACTTGGAAATATATAAAA 2188

Db 1649 GTCCCGAGACATACGCGCGAACTCAAAAGACTGTATACCCCAATTTGGAGATTAATGA 1708

QY 2189 GAAAGAAGATGATACAAACAAACCCCACTCCAGAAAAGCGGTGCTCGAAGAAG 2245

Db 1709 ACAAGACTCTGGCGAGGACAAATCCGCACATCAGTAAGCGAGCTGGCGGACGCAAGG 1765

RESULT 11

ABL28083

ID ABL28083 standard; DNA; 2010 BP.

XX AC ABL28083;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 35722.

XX KW Drosophila; developmental biology; cell signalling; insecticide;

XX KW pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Claim 1; SEQ ID NO 35722; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2010 BP; 410 A; 590 C; 571 G; 439 T; 0 other;

Query Match 1.6%; Score 59.8; DB 23; Length 2010;

Best Local Similarity 43.8%; Pred. No. 2.1e-05;

Matches 445; Conservative 0; Mismatches 542; Indels 30; Gaps 3;

QY 1229 TCAGAGAGATAAGTATTTACGACTTGGCCATCATCTCCCTTCCAAAGCCCTGTGTATGCTGC 1288

779 TCAATGTGCGCGCTGCTCGCCCTCTGGTGGCCATCGTACTGGCGCTGCATCTGT 838

QY 1289 TCGACTTCGTTAGCATGTGGTGGTGTACACTTTGGCAAAGCAAGACATCCGGGCGAT 1348

Db 839 GGTGATTTGCTCTGGGGTGATTTGTTCCGGCAGCGAAAGTGCAAGGCCATTGCGTGTG 898

QY 1349 CGGGCCTTATCTGTGGAAAGACCTTTTGGATCTGCTCTTATCTTCCAGTTG 1408

Db 899 GTATGTGGAGTGTGCTGGAGAGTACTGCTGGGCATTTGTTTACTTTATGTCATCTGTG 958

QY 1409 TTATTTGTACTTTGAGCCAGCACATTTTCGCTGTATTTCTCTAAGATGGGCTGCTTTC 1468

Db 959 CGTCCATTTCTCCCGCATCCAGGAGCGTCTCTTGGAGCCCTGGCTTCGGGAGC 1018

QY 1469 TCGGTTTTCGCTACTGTTTACGAACTGTCTACTCTCAAACTTCACAGGGTTTTGAAGGTG 1528

Db 1019 TGGGCTTTCATCACTGCTACGGCGCCATCATCTAAGTTGTACCCGCCACCTGGTGACT 1078

QY 1529 TTCTTTTCAGAACGGCTCAACGAATTCATATATGACTGGCGAGCGGTTCATGAGATGC 1588

Db 1079 TTGCGACCCGGAAGCGCATCG---CTGGGTGCTGGCGAGCTGGACCTGCTCAAGTATC 1135

QY 1589 TGGCAGTAACTACTTGGTAGTGTGTTTGGTTCTCATTTGGCTGGAGTTCATCTGTGTGCC 1648

Db 1136 TGGGCACCATGCTTTGGCGTCACTGTCTATAGCCGCCCTTACGGCTCGTCTGTGG 1195

QY 1649 AGAATTTGGAGAAACAGATTTCACTTATTTGGCCAGGGGAAAACATCCGATCACCTCATCT 1708

Db 1196 ATCTTCTCGA-----AAGTGGCGAGCTGGAGAGCCCTCAGGGAGCGGAC 1240

QY 1709 TCAATATGTGCTCATGTAGCCGCTGGGACTACATGACACAGTGTGCTGAATTTTATTC 1768

Db 1241 CGAACACCTTGCCATCCGCTCAAGTGGAACTGGTCAACGAGACACCGAGATGCTCATCC 1300

QY 1769 TCTTGTGGGGTGTATATCTCTGCTATGCACTGGGACAGTCCCATCGGCAATTCATGAGC 1828

Db 1301 TGTGCTTTGGAGTGCACCTGTCCATCGCCAGTCGGAAATGCCAACACCCAGTTCGGGAGC 1360

QY 1829 CCGCTATATGCGTGTGGAGTTCACAATGAGCTCATCATCTCTGCTATATTTCCATACAA 1888

Db 1361 GACAAATTTCTGTCACCGCACTGACGCTGGAGTTCCTGGTCTCGTCCAGCTTCTACTTTC 1420

QY 1889 TTAGATTTGCTTCTGCTCAAGACTTCATCTGATTTGGATGTTGATGTTTATTTGAC 1948

Db 1421 TGGCTTTTCTACTACCTGCGCGAAATGAGTCCCGACGCCCATCTGTGCTGGCCCTGTTTCATCC 1480

QY 1949 ATACTCATTTGACTGTGACAGTCAACATTTGGTGTGCTTTTGTATTTCCAAAGTTTTCACATT 2008

Db 1481 GCTCCCGAGCTGAGGAATAGTTTTCGGCTTGGGCTGTGATATTTGGCCAAAGTTGT----- 1534

QY 2009 CAAGCAATACCCACGAGATGATTTGCTTACAGAAGCATATGAGGATGAGCTAGACATGG 2068



Pf	30-DEC-1999;	99WQ-US311273.
Xx		
Pr	31-DEC-1998;	98US-0114495.
Xr	01-SEP-1999;	99US-0152195.
Px	(CHIR ) CHIRON CORP.	
Pa		
Pi	Barnett S, Zur Megede J;	
Dl	WIPI; 2000-452401/39.	
Dr		
Xt		
Pt	Polynucleotide encoding antigenic type C HIV gag polypeptide or a HIV	
Pt	Env polypeptide and the polypeptide useful for immunizing a mammal	
Pt	especially human against HIV	
Xt		
Ps	Claim 16; Page 97; 113pp; English.	
Xs		
Cc	Expression cassettes comprising a polynucleotide encoding antigenic	
Cc	type C human immunodeficiency virus (HIV) gag or Env polypeptides are	
Cc	used in DNA immunization, generation of packaging cell lines and	
Cc	production of Gag- and/or Env-containing proteins. Synthetic Env and	
Cc	expression cassettes exhibit increased potency for induction of	
Cc	cytotoxic T-lymphocyte (CTL) responses by DNA immunization. Gag of HIV	
Cc	self-assemble into non-infectious virus-like particles which are used	
Cc	as a matrix for the proper presentation of an antigen entrapped or	
Cc	associated to the immune system of the host.	
Xx		
Sq	Sequence 1035 BP; 189 A; 368 C; 327 G; 151 T; 0 other;	
	Query Match	1.4%; Score 52; DB 21; Length 1035;
	Best Local Similarity	45.0%; Pred. No. 0.0019;
	Matches 196; Conservative	0; Mismatches 240; Indels 0; Gaps
Qy	359	CTCTTTGCACCGGGCCTGGACACACTGACACAGCCCACTTCCTCAAGTGATGC 418
Db	347	CTGTGATGCACTGGGACCGGGAGATCAACAACACTACACCGCACCATCTACCGCCTGCTGG 418
Qy	419	TGCAGACAATAAGTCCGGGAGCAGAACAATTGACAGACGACCTGGATTGGTACCAGSGCG 478
Db	407	AGGAGACCGAACCAGCAGAGAGAACGAGAGGACCTGCTGGCCCTGGACAGCTGGC 466
Qy	479	TGTTGTGGAGCCCTTCTGGAGGGGAGCCCCAGCATCTCCGGGGGCCCATCACTTCAGCA 538
Db	467	AGAACTGTGGAACTGTTTCAGCATCACCAACTGGCTGTGGTACATCAAGATCTTCATCA 536
Qy	539	CCGATTGCTGTCCGACACGGCCCCACAGGTCTTCTCCAGGCCACCGCGGAGAGACC 598
Db	527	TGATCTGGGGGGCCTGATCGGCCCTGCGCATCATCTTCGCCCTGCTGAGCATCGTGAAC 596
Qy	599	GCATCTCTGCTCCAAGACCTGTCTCTCCGACCCACCCAGCTGGCCACAGCCCACTCTGGAGA 658
Db	587	CGTGTGCCAGGGGTACAGCCCCTTGCCTTCAGACCCCTGACCCCAACCCCGCGAGC 646
Qy	659	CCGAGTGTGTTCCACGGGCTCCGGCGCAAGTGGAGGGCCCCACTTACACCGCGCGGCCCA 718
Db	647	CCGACCCCTTGGGCGCATCGAGAGGAGGGCGGCGAGCAGGACCGCGCGCAGCATCC 706
Qy	719	ATCAGGGGCCCCGGGCTCTGGGCACAGCTGGGGGGCGCAGGACGGCTCGGGGGGACA 778
Db	707	GCCTGTGTGAGCGGCTTCTTGGCCCTGGCCCTGGGACGACCTGCGCAGCCTGTGCTGTTCA 766
Qy	779	AGAGCCACTTCAAGTG 794
Db	767	GCTACCAACCGCTGCG 782
	RESULT 14	
	AAA51614	
ID	AAA51614 standard; DNA; 2466 BP.	
Xx		
Ac	AAA51614;	
Xx		
Xx	31-OCT-2000 (first entry)	









GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 15, 2003, 04:20:23 ; Search time 4738 Seconds  
(without alignments)  
12469.636 Million cell updates/sec

Title: US-09-775-181-1  
Perfect score: 3648  
Sequence: 1 atggagccatggcttacc.....gggatagttttaagttag 3648

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Maximum Match 100%  
Listing first 45 summaries

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :	EST:*
1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hic:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_hic:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_vit:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	535.6	14.7	680	10	BB429778
2	502.8	13.8	628	10	BB644412
3	439.2	12.0	628	14	BQ109073
4	380.8	10.4	508	17	AQ764784
5	372.4	10.2	554	17	AQ837108
6	336.2	9.2	543	17	AQ728799

c	7	222.6	6.1	681	9	AU171941	AU171941
c	8	219.2	6.0	280	10	BB354926	BB354926
c	9	172.6	4.7	895	17	CNS032UA	AL225307 Tetraodon
c	10	170.4	4.7	356	10	BB872944	BB872944
c	11	168.8	4.6	755	10	BB255902	BB255902
c	12	166.8	4.6	632	17	CNS02NOT	AL205670 Tetraodon
c	13	165	4.5	961	17	CNS032LV	AL178348 Tetraodon
c	14	148.4	4.1	358	10	BB68702	BB68702
c	15	108.2	3.0	592	17	AQ881884	AQ881884 HS_5304_A
c	16	108	3.0	382	14	BQ027934	BQ027934 UI-H-CO0-
c	17	101.2	2.8	877	17	CNS03H19	AL243702 Tetraodon
c	18	95.4	2.6	329	17	AZ628814	AZ628814 LM048D17
c	19	93.2	2.6	1030	17	CNS03EM8	AL240569 Tetraodon
c	20	90.4	2.5	610	10	BB621973	BB621973
c	21	88.6	2.4	740	10	BB025325	BB025325
c	22	87.2	2.4	988	17	CNS04BT9	AL283590 Tetraodon
c	23	86	2.4	633	17	CNS04FKJ	AL288460 Tetraodon
c	24	86	2.4	921	17	CNS03PB5	AL254426 Tetraodon
c	25	85.6	2.3	686	10	BB618140	BB618140
c	26	84.8	2.3	703	17	BB323489	BB323489 CH230-197
c	27	82	2.2	888	17	CNS03YL7	AL266452 Tetraodon
c	28	80.8	2.2	807	17	CNS0317B	AL223184 Tetraodon
c	29	80.6	2.2	957	17	CNS020UA	AL207163 Tetraodon
c	30	80.4	2.2	822	17	CNS02FTA	AL195463 Tetraodon
c	31	80	2.2	686	10	BB027806	BB027806
c	32	79.8	2.2	655	10	BB632654	BB632654
c	33	78.2	2.1	899	17	CNS032UB	AL225308 Tetraodon
c	34	78	2.1	483	17	AZ903235	AZ903235 RPCI-24-1
c	35	77	2.1	1096	17	CNS05CMB	AL331292 Tetraodon
c	36	76.8	2.1	616	10	BB178309	BB178309
c	37	75.6	2.1	379	17	CNS03EM7	AL240568 Tetraodon
c	38	73.2	2.0	1012	17	CNS03Y7G	AL265957 Tetraodon
c	39	72.8	2.0	446	17	AZ870959	AZ870959 2M0183F08
c	40	65	1.8	563	17	BB315466	BB315466 CH230-196
c	41	65	1.8	805	12	BG687462	BG687462 602639433
c	42	63.6	1.7	817	17	CNS05ATS	AL328969 Tetraodon
c	43	63.4	1.7	925	17	CNS0091P	AL053013 Drosophil
c	44	61.8	1.7	591	17	AQ764779	AQ764779 HS_3118_A
c	45	61.6	1.7	487	12	BG813029	BG813029 daf70c02.

ALIGNMENTS

RESULT 1  
BB429778  
LOCUS BB429778 680 bp mRNA linear EST 25-OCT-2001  
DEFINITION BB429778 RIKEN full-length enriched, adult male hippocampus Mus musculus cDNA clone C630030A14 3', mRNA sequence.

ACCESSION BB429778  
VERSION BB429778.2 GI:16424024

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Arkawala, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 680)

TITLE RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
JOURNAL Unpublished (2001)  
COMMENT On Jul 18, 2000 this sequence version replaced gi:9269505.

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222



Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Itoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
source

Location/Qualifiers  
1. .628  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B230315M17"  
/clone\_lib="RIKEN full-length enriched, adult male corpora quadrigenina"  
/sex="male"  
/tissue\_type="corpora quadrigenina"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATTAATTCCTCCCTCCCTCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 110 a 176 c 222 g 119 t 1 others

ORIGIN

Query Match 13.8%; Score 502.8; DB 10; Length 628;  
Best Local Similarity 87.6%; Pred. No. 8.2e-112;  
Matches 549; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 116 CCCCAGAGGGGAGCGCAGCCAGCCAGCGGTGTCAGCCTCTGCTCGGACTCTCTCGG 175  
DB 628 CTCNGAGGGGAGCAGCATGTGTCAGCAGCTGTCTGCCAGACCCCTCTTA 569

QY 176 CTCCTTGGAGCGCTCCACCGATGGCACCATCTTGGCGAGAACTCTGCCGAGGAGGTGC 235  
DB 568 TTCCTTGGAGCGCTCCATCTGACGGCACCATCTTGGCAGAAAACCTAGCCGAGGAGGTGC 509

QY 236 CCATGACGCTGGCCTTACCTCTACACGGGGACTCTCCACCACTGTAAGCGAGCAACT 295  
DB 508 CCGTGACGCTGGCCTTACCTCTACACTGGGACTTCCACCACTGTAAGAGCAACT 449

QY 296 GCTCCGGCGCTACGAGTTGGCGGCGCTCCCGGGGAAGTGGCCAGCCCTGGCCAGCGCGC 355  
DB 448 GCTCAGGCGCTACGAGTTGGCTGGCTCCCGGAAAGTCACCGTCCCTAGCCAGCTCC 389

QY 356 ACCCTCTCTTGACCGGGCGCTGGACACACTGACACAGCCACCACTTCTTCAACGTGA 415  
DB 388 ATCTCTCTTGACCGGGCGCTGGACACTCTGACACAGCCACCACTTCTTCAACATGA 329

QY 416 TGCTGAGAGCAATAAGTCGCGGAGCAGAACTTGACGAGCAGCACTGGATTGGTACCAGG 475  
DB 328 TGTGTCAGAGCAATAAGTCGCGGAGCAGCACTGACGATGACCTGCACTGAGTGTACCAGG 269

QY 476 CGCTGGTGTGGAGCCTCTTGGAGGGAGCCAGCAGCATCTCCCGGGGGCCATCACCTTCA 535  
DB 268 CTCTGGTGGAAAGTCTCTTGGAGGGTGAAGCCAGCATCTCCCGGGGGCCATCACCTTCA 209

QY 536 GCACCGATTTCGTCTCCGACCGGCCACACAGGTCTTCTCCAGGCGACGCGGAGGAGA 595  
DB 208 GCACCGAGTCTGTCTGACGCGGCCCGCCGACGAGTCTTCTCCAGGCGACCCGAGGAGA 149

QY 596 GCGCATCTGTCTCAAGACCTGTCTCTCCGCGACCCACCTGGCCAAAGCACTCTGG 655  
DB 148 GTCGCATCTGTCTGACGAGCACTGTCTCTCTCGCTCACCACCTGGCTAAGCGCACTTAG 89

QY 656 AGACCGAGTGTCTTCCACGCGCTCCGCGCAAGTGGAGGCGCCCACTTACAGCGCGCGCGC 715  
DB 88 AGACCGAGTGTCTTCCATGGGCTTCGCGGAAAGTGGAGGCGCCCACTTGCACCGCGGTGCT 29

QY 716 CCAATCAGGGGCCCGCGGCGCTGGGCC 742  
DB 28 CCAATCAGGGGCCACGCGGCGCTAGGCC 2

RESULT 3  
B0109073  
LOCUS  
DEFINITION  
IMAGE:q8\_2001/sno349bdr81.yl NIH\_MGC\_94 Mus musculus cDNA clone  
IMAGE:5365177 5', mRNA sequence.  
ACCESSION  
B0109073  
VERSION  
B0109073.1  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 628)  
Kale, P.I., Harsch, T.J., Folta, P.A., Nelson, D.O., Sanders, C.G. and Prange, C.K.  
The I.M.A.G.E. Consortium quality control effort: clone  
ressequencing for verification  
Unpublished (2001)  
Contact: Prange CK  
The I.M.A.G.E. Consortium  
Lawrence Livermore National Laboratory  
Livermore, CA, USA  
Email: [help@image.llnl.gov](mailto:help@image.llnl.gov)  
This read has been produced as part of the I.M.A.G.E. Consortium  
quality control effort. High quality sequence is defined as having  
100 or more base pairs with a phred quality value of 20 or greater,  
where a sliding window of 4 base pairs with a phred quality value  
of 15 or greater marks the beginning and end of the sequence. For  
information on obtaining this clone, please contact  
[info@image.llnl.gov](mailto:info@image.llnl.gov).  
Plate: LLAM11929 row: c column: 2  
Seq primer: ml3rpl  
High quality sequence stop: 628.  
Location/Qualifiers  
1. .628  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5365177"  
/clone\_lib="NIH\_MGC\_94"  
/tissue\_type="retina"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: eye; Vector: pCMV-Sport6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.

FEATURES  
source



AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999), 99380589

CONTACT: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

(pieter@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu

Plate: 1040 row: O column: 3

Seq primer: T7

Class: BAC ends

High quality sequence stop: 554.

Location/Qualifiers

1. .554

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=1040 Col=3 Row=O"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 182 a 147 c 119 g 102 t 4 others

ORIGIN

Query Match 10.2%; Score 372.4; DB 17; Length 554;

Best Local Similarity 92.0%; Pred. No. 6e-80;

Matches 414; Conservative 0; Mismatches 33; Indels 3; Gaps 2;

QY 2157 AAACTCTATGCCCACTGGAATATATAAAGAAAGAGATGATCACAACACCCCA 2216

Db 93 AAACTCTATGCCCACTGGAATATATAAAGAAAGAGATGATCACAACACCCCA 152

QY 2217 CTCCAGAAAAAGCGGTGCTCGAAGAGGGCCTAGTCTTCATCATGAGAGCATTTAC 2276

Db 153 CTCCAGAAAAAGCGGTGCTCGAAGAGGGCCTANGTCCGTCCATCATGAGAGCATTTAC 212

QY 2277 GGAGATCCAGACAGTACAGCGGCGAGTCTCTAAAGAGGACAGAGGGCGCCGACCA 2336

Db 213 GGAGATCCAGACAGTACAGCGGCGAGTCTCTAAAGAGGACAGAGGGCGCCGACCA 272

QY 2337 TGGCAGAGCAAAAGGCACCTGCTCATCAGGAAGACCCCGAGAGTCTTCAGGGAACAC 2396

Db 273 TGGCAGAGCAAAAGGCACCTGCTCATCAGGAAGACCCCGAGAGTCTTCAGGGAACAC 332

QY 2397 AGGGAATCCAGAGGAGACCCCTGAAAAACCGAGTCTTCTACTCAAGAAATCCACAG 2456

Db 333 AGGGAATCCAGAGGAGACCCCTGAAAAACCGAGTCTTCTACTCAAGAAATCCACAG 392

QY 2457 CACTTATGACCAGTCCAGAGACCAACGAGAGTCCAGTAGCTACCACAGAAAGCCA 2516

Db 393 CACTTATGACCAGTCCAGAGACCAACGAGAGTCCAGTAGCTACCACAGAAATGTC- 451

QY 2517 AGAGGAGGAGACAAAGAAAATCCACACTGGAAATCCCTGTGGGTAAAAAATAACACA 2576

Db 452 AGAGGAGGAGACTACAGAGAAATCCACACTGGAAATCCCTGTGGGTAAAAAATAACACA 509

QY 2577 AAAAATAAAGAGACAGCGAGGCTGAGTC 2606

Db 510 GCTATCTATAGAAGACAGCAGGCTTATTC 539

RESULT 6

AQ728799

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong

(pieter@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm)

or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu

Plate: 1040 row: B column: 22

Seq primer: T7

Class: BAC ends

High quality sequence stop: 543.

Location/Qualifiers

1. .543

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=1040 Col=22 Row=B"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 168 a 139 c 124 g 107 t 5 others

ORIGIN

Query Match 9.2%; Score 336.2; DB 17; Length 543;

Best Local Similarity 87.4%; Pred. No. 4.2e-71;

Matches 368; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 2146 GACGAGTGAAGAACTCTATGCCCACTGGAATATATAAAGAAAGAGATGATCACA 2205

Db 110 GACGAGTGAAGAACTCTATGCCCACTGGAATATATAAAGAAAGAGATGATCACA 169

QY 2206 ACAACCCCACTCCAGAAAAAGCGGTGCTCGAAGAGGGCCTAGTCTTCCATCATG 2265

Db 170 ACAACCCCACTCCATGATAAGCGGTGCTCGAAGAGGGCCTAGATCATG 229

QY 2266 AGACGATTACGAGATCCAGAGACAGTACGCCGCGAGTCTCTAAAGAGGACAGGAG 2325

Db 230 AGACTCATTACGAGATCCAGAGACAGACAGCGCAATGCTCTAAAGAGGACAGAG 289

QY 2326 GCGCGGACCATGGCACAGCCAAAGGCACTGCCCTCATCAGGAAAGACCCCGCAGAGTCT 2385

```

Db 290 GCGCGCACACAGTGGACACAGTCAAACTCTGTCTCTCATCAAGAGAACCCCGAGAGTCT 349
QY 2386 TCAGGGAACACAGGGAAATCCAGGAGGAGACCCCTGAAAACCGAGTCTTCTCACTCAAG 2445
Db 350 TCACGGAACACACAGGAAATACAGAGGAGAGACCCCTGAATACCGAGTCTTGTACTCTGAG 409
QY 2446 AAATCCACAGCAGTATATGACCACTGTGAGAGACCAACGAGAGTCCAGTACGCTACCC 2505
Db 410 AAATACCACAGCTTATGACCACTGTGAGAGACACACACGAGACCTCTCGTACCTACTC 469
QY 2506 ACAGAAGCCAGAGAGAGAGACACAGAAAATTCACACTGGAATCCCTGTGCGGTAA 2565
Db 470 CTCGAAAGCTGAGAGAGAGACATCAGACTATTGACACATGTGTATCCCTGTGCGAGAAA 529
QY 2566 A 2566
Db 530 A 530

RESULT 7
LOCUS AUI171941/c
DEFINITION AUI171941 681 bp mRNA linear EST 29-JAN-2001
ACCESSION AUI171941
VERSION AUI171941.1 GI:12594012
KEYWORDS EST.
SOURCE Japanese medaka.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomerygia; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 681)
Mita, K., Ishikawa, Y. and Yamauchi, M.
Establishment of cDNA database of medaka, Oryzias latipes
Unpublished (2001)
Contact: Mita K
Genome Research Group
National Institute of Radiological Sciences
Angawa 4-9-1, Inage, Chiba 263-8555, Japan
Email: kmita@nirs.go.jp
method: uni-directional sequence direction: sequenced from T3 primer
(5' -> 3').

FEATURES
Source
1..681
Location/Qualifiers
1..681
/organism="Oryzias latipes"
/strain="HNI"
/db_xref="taxon:8090"
/clone="br8300"
/clone_lib="OI-br-ad cDNA"
/sex="female/male mixed"
/tissue_type="brain"
/dev_stage="adult"
BASE COUNT 124 a 173 c 185 g 199 t
ORIGIN

Query Match 6.1%; Score 222.6; DB 9; Length 681;
Best Local Similarity 65.2%; Pred. No. 2.6e-43;
Matches 379; Conservative 0; Mismatches 189; Indels 13; Gaps 3:

QY 1921 GATTGGATTTGATGCTGTATTTTGCATACATCTATTTGACTGTGACAGTCCACATTTGGG 1980
Db 681 GACTGGATGCTATGCTGTTTTTGTCTACACTACCTGACTGTGACTGTGACTGTGGC 622
QY 1981 TTGCTTT-TTGATTCCAAAGTTTTCACATTCAGCAATTAACCCAGAGATGATATGCTAC 2039
Db 621 CTTCTGCTTTGCCAAAGTTTCCCTGTCAAAAGGCATTCAGGCAAGGATGACATCGCCAC 562
QY 2040 AGAACCATATGAGGATGAGCTAGACATGGCGCGATCTGGATCCTACCTCAACAGCATAT 2099
Db 561 AGAGGCATACGAGAGGAGCTTGACATGGGAAGATCTGGGTCTTACCTCAACAATGAT 502

```

```

QY 2100 CAATTCCAGCTGGAGTGAGCACAGCTTTGGATCCAGAGGACATTCGGGACGAGTGAATAA 2159
Db 501 CACCTCTGCTGGAGCGAGCACAGCTTGACCTGTATATACGGGATGAATTTGAAGA 442
QY 2160 ACTCTATGCCCACTGGAAATATATAAAGAAAGAGATGATCACAAACACCCACCT 2219
Db 441 GCTGTATGCACAACCTGGAAGTCTACAAAGAAAGAGATGCTCGCCACCAACCCACCT 382
QY 2220 CCAGAAAAGCGGTGCTCGAAGANGGCTAGTGGTTCATCATCATGACGAGCATACGGA 2279
Db 381 TCAAAAGAAACGCAACTCAAAAGAGGCGCTGGAGCGACCCCTGATGAACGATCACAGA 322
QY 2280 GATCCAGAGACAGTCAGCCGCGAGTCTCTAAAGAGGACAAAGAGGCGGCGGACCATGG 2339
Db 321 GATCCAGAGACCATGCACATCCACCGCAGTGCAGCCGAGAGCGGCGGAGCAGCACGG 262
QY 2340 CACAGCCAAAGGACATGCCCTCATCAGGAAGAACCCCGAGAGTCTTTCAGGGAACACAGG 2399
Db 261 CAGCAACCGGGGCAC-----CTTGAAGAGAAACAGATTTGAGCCAAAGCCACCATGG 211
QY 2400 GAAATCCAGAGGAGAGACCCCTGAAAACCGAGTCTTCTCACTCAAGAAATCCACAGCAC 2459
Db 210 AAAACCTCGGGATGACTCCCTTTAAAACAAAGTCTATGTTTGGCGAAGTCACTCAGC-- 153
QY 2460 TTATGACCACGTGAGAGACCAAAACGGAAGAGTCCAGTAGCC 2500
Db 152 -TTTGACCACGTTTGTGACCAGGATGCAGAAACTGGGAACC 113

RESULT 8
LOCUS BB354926
DEFINITION BB354926 280 bp mRNA linear EST 12-JUL-2000
ACCESSION BB354926
VERSION BB354926.1 GI:9066754
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 280)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci
P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N.,
Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata
Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y.,
Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya
T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamanaka, I.,
Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino
M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9216
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermotabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kikunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
Y. and Hayashizaki, Y.

```





```

Db 483 AGTGACCCAGGTCGGTCTCTGCTGCAAGTCGGCCAGCGCTCATACCTCTCGGGG 424
QY 2663 ACAAGAAACTGGGACCCACCAACATCGATCTACAGAAGTCTCTCAGTGFCATAGCAA 2722
Db 423 ACAAGAAGCCATTCCTCCGACGAGGTCATCTCGAAGAGTCCCTCAGCGTCATCGCTG 364
QY 2723 GGCCCAAGGAGAGACTCTTGATAGCTGGGAAACCCCAAC 2765
Db 363 GCGCCAAAGAGAGGAGCGCTCGGGTTGACGCCAAAGGCCAGAAC 321

RESULT 10
BB872944
LOCUS BB872944 356 bp mRNA linear EST 27-NOV-2001
DEFINITION BB872944 RIKEN full-length enriched, 16 days neonate male
diencephalon Mus musculus cDNA clone G630041A19 5', mRNA sequence.
ACCESSION BB872944
VERSION BB872944.1 GI:17119154
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 356)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,K., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
,A., Takahashi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
JOURNAL
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Prepare full-length cDNA libraries for cap-trapper-selected cDNAs to
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
FEATURES
source
Location/Qualifiers
1..356
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G630041A19"
/clone_lib="RIKEN full-length enriched, 16 days neonate
male diencephalon"
/sex="male"

```

```

/tissue_type="diencephalon"
/dev_stage="16 days neonate"
/note="pooled tissues; (tissue_type=cerebellum,
dev_stage=16 days neonate, sex=mixed),
(tissue_type=cerebellum, dev_stage=0 day neonate,
sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
sex=male), (tissue_type=whole body, dev_stage=9 days
embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days
embryo, sex=mixed)"
BASE COUNT 113 a 83 c 82 g 78 t
ORIGIN

Query Match 4.7%; Score 170.4; DB 10; Length 356;
Best Local Similarity 82.6%; Pred. No. 1.3e-30;
Matches 195; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 3412 GGACACCAGGAAAAAGACATCTTCTCTGAGGAGATGTCGGTCTCTATTAACCTCA 3471
Db 2 GGAGACCAGGAGAAACAGACATCTTCTCTGTGGATATCATCTCTGGCTCTGTAACCTCA 61
QY 3472 AGTAATAACTTCCAGCAACCTTTACATCAGCAGCAGAGGTTTGTCCTTGGAGTTTCAG 3531
Db 62 AGTAATAACTCCATCAGCCCTTAACATCAGCAGCGGAGTGTGTCCTGGAGTTCAG 121
QY 3532 ACCCAGCTCAACCAAAATGCTGGAGAGTGTAGCTTTTACCTGCTCTCTCTCTAAGT 3591
Db 122 CCCCTAGAACCAACCAATGCTGAAAGAGCGTAACTTTACCTGCTCTCTCTTAAAGT 181
QY 3592 GCAATAAGATGAGCGCCCTAGGAAAGAGATCTCGGATAGTTTAAAGTGTA 3647
Db 182 GCAATAAGATACCAAGGCGCTCAGAAATAGGAAGTCTGGACACTTTCAGAAAGTGA 237

RESULT 11
BB255902
LOCUS BB255902 755 bp mRNA linear EST 24-OCT-2001
DEFINITION BB255902 RIKEN full-length enriched, 7 days neonate cerebellum Mus
musculus cDNA clone A730070K20 3', mRNA sequence.
ACCESSION BB255902
VERSION BB255902.2 GI:16398506
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 755)
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
On Jul 6, 2000 this sequence version replaced gi:8948735.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and

```





URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.  
e mouse tissues.

## FEATURES

Location/Qualifiers

1..358  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="G630007J06"  
/clone.lib="RIKEN full-length enriched, pooled tissues,  
intestinal mucosa, etc."  
/note="pooled tissues : (tissue\_type=intestinal mucosa,  
dev\_stage=adult, sex=male), (tissue\_type=accessory  
axillary lymph node, dev\_stage=adult, sex=male),  
(tissue\_type=jejunal and colic lymph node, dev\_stage=adult  
, sex=male), (tissue\_type=gall bladder, dev\_stage=adult,  
sex=male), (tissue\_type=vesicular gland, dev\_stage=adult,  
sex=male), (tissue\_type=spinal cord, dev\_stage=11 days  
embryo), (tissue\_type=brain, dev\_stage=13 days embryo),  
(tissue\_type=lung, dev\_stage=14 days embryo),  
(tissue\_type=ovary and uterus, dev\_stage=15 days embryo),  
adult, sex=female), (tissue\_type=cortex, dev\_stage=0 day  
neonate), (tissue\_type=cerebellum, dev\_stage=1 month  
neonate), (tissue\_type=diaphragm, dev\_stage=16 days  
neonate, sex=male), (tissue\_type=medulla oblongata,  
dev\_stage=16 days neonate, sex=male),  
(tissue\_type=cerebellum, dev\_stage=21 days neonate),  
(tissue\_type=testis, dev\_stage=8 days neonate, sex=male)"

BASE COUNT

113 a 84 c 83 g 78 t

## ORIGIN

Query Match 4.1%; Score 148.4; DB 10; Length 358;  
Best Local Similarity 81.9%; Pred. No. 3.1e-25;  
Matches 195; Conservative 0; Mismatches 41; Indels 2; Gaps 2;  
QY 3412 GGACACCGAGAAAAAGACATCTT-CTTCTGAGGAGATGTGCGTGGCTCCTATAACTC 3470  
Db 2 GGAGACCGAGAAACAGACATCTTCTCTGTGGATATATCTCTGGCTGCTGAACCTC 61  
QY 3471 AAGTATATCTTCCAGCAACCTTTACATCAGCAGCAGAGGTTTGCTCTGGGAG-TTTG 3529  
Db 62 AAGTATATCTTCCATCAGCCCTTAAACATCAGCAGCGAAGTGTGTCTCTGGGAGCTCG 121  
QY 3530 AGACCCAGCTCAACCAAAATGCTGGAAGAGTGTAGCTTTACCTGCCTCTCTCTCTAA 3589  
Db 122 AGCCCTAGAACACCAAAATGCTGAAGAGCGTAACTTACCTGCCTCTCTCTCTAA 181  
QY 3590 GTGCAATAAGATAGCAGGCGCTAGGAAGAAGAGATCTGGGATAGTGTTTAAAGTGA 3647  
Db 182 GTGCAATAAGATACACGCGCTCAGAAATAGGAAGTCTGGCACACTTTCAGAGTGA 239

RESULT 15

AQ881884

## LOCUS

## DEFINITION

AQ881884  
HS\_5304\_ALG04\_SP6E\_RPCI-11 Human Male BAC Library GSS 09-NOV-1999  
genomic clone Plate-9072 Col=7 Row=M, DNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AQ881884.1

GI:6313351

GSS.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 592)

Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,

Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and

Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

99380589

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm))or from Research h Genetics ([inforesgen.com](http://inforesgen.com)). BAC end Web Server:<http://www.htsc.washington.edu>

plate: 9072 row: M column: 7

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 592.

Location/Qualifiers

1..592

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate-9072 Col=7 Row=M"

/clone.lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;

Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and

EcoRI methylase. Size selected DNA was cloned into the

pBACE3.6 vector at EcoRI sites"

166 a 94 c 97 g 219 t 16 others

BASE COUNT

ORIGIN

Query Match 3.0%; Score 108.2; DB 17; Length 592;

Best Local Similarity 96.5%; Pred. No. 2.3e-15;

Matches 110; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1889 TTAGATTGTTCTTGCCTCAAGACTTCAGTCTGATTGGATGTTGATGCTGTTATTTGCAC 1948

Db 81 TCAGATTGTTCTTGCCTCAAGACTTCAGTCTGATTGGATGTTGATGCTGTTATTTGCAC 140

QY 1949 ATACTCATTTGACTGTGACAGTCACCATTTGGCTTCTTTTGTATCCAAAGTTT 2002

Db 141 ATACTCATTTGACTGTGACAGTCACCATTTGGCTTCTTTTGTATCCAAAGTTAT 194

Search completed: February 15, 2003, 05:54:58

Job time : 4773 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2003, 22:48:02 ; Search time 115 seconds  
(without alignments)  
9728.320 Million cell updates/sec

Title: US-09-775-181-1  
Perfect score: 3648  
Sequence: 1 atggagccatggtacc.....gggtagtttaaaagttag 3648

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2.6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	1.7	7218	1	US-08-232-463-14
2	48.4	1.3	7218	1	US-08-232-463-14
3	46.4	1.3	2481	1	US-08-324-243-35
4	46.4	1.3	2481	1	US-08-532-390-35
5	46.4	1.3	2481	3	US-08-717-294-35
6	46.4	1.3	2481	5	PCT-US95-11511-35
7	44.8	1.2	2502	1	US-08-073-384C-7
8	44.8	1.2	2502	1	US-08-254-359A-7
9	44.8	1.2	2502	1	US-08-483-043-7
10	44.8	1.2	2502	1	US-08-481-238-7
11	44.8	1.2	2502	2	US-08-471-066B-7
12	44.8	1.2	2502	2	US-08-484-956-7
13	44.8	1.2	2502	2	US-08-757-653-7
14	44.8	1.2	2502	2	US-08-599-491-7
15	44.8	1.2	2502	2	US-08-756-386-7
16	44.8	1.2	2502	2	US-08-823-516-7
17	44.8	1.2	2502	3	US-08-682-853A-7
18	44.8	1.2	2502	3	US-08-759-038-7
19	44.8	1.2	2502	3	US-08-758-314-7
20	44.8	1.2	2502	4	US-09-350-309-7
21	44.8	1.2	2502	4	US-08-520-946-7
22	44.8	1.2	2634	3	US-08-911-853-30
23	44.8	1.2	2634	4	US-09-479-409-30
24	44.8	1.2	2634	4	US-09-479-409-30
25	44.8	1.2	17612	3	US-08-911-853-29
26	44.8	1.2	17612	4	US-09-479-409-29
27	44.8	1.2	17612	4	US-09-479-453-29

c 28	44.4	1.2	2495	4	US-09-141-027-1	Sequence 1, Appli
c 29	42.8	1.2	771	2	US-08-972-008-3	Sequence 3, Appli
c 30	42.8	1.2	771	4	US-09-267-409-3	Sequence 3, Appli
c 31	42.8	1.2	2525	2	US-08-972-008-1	Sequence 1, Appli
c 32	42.8	1.2	2525	4	US-09-267-409-1	Sequence 1, Appli
c 33	42	1.2	5382	4	US-09-479-122-21	Sequence 21, Appli
c 34	42	1.2	9737	4	US-09-479-122-22	Sequence 23, Appli
c 35	42	1.2	9737	4	US-09-479-122-23	Sequence 23, Appli
c 36	42	1.2	9737	4	US-09-479-122-28	Sequence 28, Appli
c 37	42	1.2	9871	4	US-09-479-122-24	Sequence 24, Appli
c 38	42	1.2	10060	4	US-09-479-122-25	Sequence 25, Appli
c 39	42	1.2	44377	2	US-08-804-227C-7	Sequence 7, Appli
c 40	42	1.2	44377	2	US-08-804-198-1	Sequence 1, Appli
c 41	41.8	1.1	2496	1	US-08-073-384C-2	Sequence 2, Appli
c 42	41.8	1.1	2496	1	US-08-254-359A-2	Sequence 2, Appli
c 43	41.8	1.1	2496	1	US-08-483-043-2	Sequence 2, Appli
c 44	41.8	1.1	2496	1	US-08-481-238-2	Sequence 2, Appli
c 45	41.8	1.1	2496	2	US-08-471-066B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-fls  
US-08-232-463-14

Query Match 1.7% Score 62; DB 1; Length 7218;





TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2481 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-324-243-35

Query Match 1.3%; Score 46.4; DB 1; Length 2481;  
Best Local Similarity 44.9%; Pred. No. 0.022;  
Matches 176; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 359 CTTCTTGCACCGGGCTGGACACTGACACAGCCACCACTTCTTCAACGTGATGC 418  
DB 1796 CTTGATGAGTGGGAGCGCGAGATCGATACTACACAGCCTGATCTACAGCCTGCTGG 1855  
419 TGCAGAGCAATAAGTCGCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACCGAGCGC 478  
1856 AGAAGAGCCAGACCAGCAG 1915  
QY 479 TGGTGTGGAGCCTTCTTGGAGGGCGGAGCCAGCATCTCCCGGGCGGCATCATCTTCAGCA 538  
DB 1916 CGAGCCTGTGGAACCTGTTTCGACATCACCAACTGGCTGTGGTACATCAAAATCTTCATCA 1975  
QY 539 CCGATTTCGCTGCCAGCCGCGCCACAGGTCTTCTCCGACACCCACCTGGCCAAAGCCACTCTGGAGA 598  
DB 1976 TGATTGTGGGGCTGTGGGCGCTCCGCGATCGTGTTCGCCGTGTGAGCATCTGTAACC 2035  
QY 599 GCATCTGCTTCCAAAGACCTGTCTCTCCGACACCCACCTGGCCAAAGCCACTCTGGAGA 658  
DB 2036 GCGTGGCGGAGGCTACAGCCCTCAGGCTCCAGAGCCGCGCCCGCTGGCGCGGGC 2095  
QY 659 CCGAGTGTTCACAGGCTTCCGCGCAAGTGGAGGCGCCACATTCACAGCCGCGGGCGCCA 718  
DB 2096 CGGACCGCGCGGAGGCTACGAGGAGGGGGGGGAGGCGGAGGCGGAGGCGGAGGCGGCA 2155  
QY 719 ATCAGGGGCGCGGGCTGGCCACAGCTGG 750  
DB 2156 GCGTGTGCACGGCTTCTTGGCGATCATCTGG 2187

RESULT 4  
US-08-532-390-35  
Sequence 35, Application US/08532390  
Patent No. 5795737  
GENERAL INFORMATION:  
APPLICANT: SEED, BRIAN  
APPLICANT: HAAS, JURGEN  
TITLE OF INVENTION: High Level Expression of Proteins  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/532,390  
FILING DATE: 22-SEP-1995  
PRIOR APPLICATION NUMBER: 08/324,243  
FILING DATE: 19-SEP-1994  
ATTORNEY/AGENT INFORMATION:

NAME: LECH, KAREN F.  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/294001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2481 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-532-390-35

Query Match 1.3%; Score 46.4; DB 1; Length 2481;  
Best Local Similarity 44.9%; Pred. No. 0.022;  
Matches 176; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 359 CTTCTTGCACCGGGCTGGACACTGACACAGCCACCACTTCTTCAACGTGATGC 418  
DB 1796 CTTGATGAGTGGGAGCGCGAGATCGATACTACACAGCCTGATCTACAGCCTGCTGG 1855  
QY 419 TGCAGAGCAATAAGTCGCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACCGAGCGC 478  
DB 1856 AGAAGAGCCAGACCAGCAG 1915  
QY 479 TGGTGTGGAGCCTTCTTGGAGGGCGGAGCCAGCATCTCCCGGGCGGCATCATCTTCAGCA 538  
DB 1916 CGAGCCTGTGGAACCTGTTTCGACATCACCAACTGGCTGTGGTACATCAAAATCTTCATCA 1975  
QY 539 CCGATTTCGCTGCCAGCCGCGCCACAGGTCTTCTCCAGGCGCCAGCGCGAGGAGCGC 598  
DB 1976 TGATTGTGGGGCTGTGGGCGCTCCGCGATCGTGTTCGCCGTGTGAGCATCTGTAACC 2035  
QY 599 GCATCTGCTTCCAAAGACCTGTCTCTCCGACACCCACCTGGCCAAAGCCACTCTGGAGA 658  
DB 2036 GCGTGGCGGAGGCTACAGCCCTCAGGCTCCAGAGCCGCGCCCGCTGGCGCGGGC 2095  
QY 659 CCGAGTGTTCACAGGCTTCCGCGCAAGTGGAGGCGCCACATTCACAGCCGCGGGCGCCA 718  
DB 2096 CGGACCGCGCGGAGGCTACGAGGAGGGGGGGGAGGCGGAGGCGGAGGCGGAGGCGGCA 2155  
QY 719 ATCAGGGGCGCGGGCTGGCCACAGCTGG 750  
DB 2156 GCGTGTGCACGGCTTCTTGGCGATCATCTGG 2187

RESULT 5  
US-08-717-294-35  
Sequence 35, Application US/08717294  
Patent No. 6114148  
GENERAL INFORMATION:  
APPLICANT: SEED, BRIAN  
APPLICANT: HAAS, JURGEN  
TITLE OF INVENTION: HIGH LEVEL EXPRESSION OF  
PROTEINS  
NUMBER OF SEQUENCES: 110  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Clark & Elbing LLP  
STREET: 176 Federal Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/717,294  
FILING DATE: 20-SEP-1996



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/073,384C  
FILING DATE: 04-JUN-1993  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/986,330  
FILING DATE: 07-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: FORS-00613  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/705-8410  
TELEFAX: 415/397-8338  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2502 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-073-384C-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;  
Best Local Similarity 46.8%; Pred. No. 0.06;  
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;  
QY 413 TGATGCTGAGAGCAATAAGTCGCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACC 472  
DB 1820 TGGTGGCCCTGGACTATAGCCAGATAGACTCCGGGTCTCTGCCACCTCTCCGGGGAGG 1879  
QY 473 AGCGCTGTGTGGAGCCTTCGAGGGGAGCGGAGCAGCATCTCCGGGGGCGCATCACT 532  
DB 1880 AGAAGCTGATCGGTCTTCCAGGAGGGAGGAGCATCCACACCGAGCGGCGCAAGACCATCA 1939  
QY 533 TCAGCAGCAGATTGCTGTCCGACCGCGCCACAGGTCTTCTCCAGGCGCAGCGCGAGG 592  
DB 1940 TGTTCGGCTCCCGGAGGCGGTGGACCCCTGATGGCGGGCGGCGCAAGACCATCA 1999  
QY 593 AGAGCGGATCTGCTCCAGACCTGTCTCTCCGACCGCCACCTGGCCAGCGGCGCATC 652  
DB 2000 ACTTCGGGTC---CTCTACGGCATGTCCGCCACCGCTCTCCAGGAGCTTGCATCC 2056  
QY 653 TGGAGACCGAGTGGTTCACGCGCTCCGCGCAAGTGGAGGCGCCACTTACACCGCGG 712  
DB 2057 CCTACGAGGCGGTGGCTTCTTATGAGCGCTACTTCCAGAGCTTCCCAAGGTGCGGG 2116  
QY 713 GCCCAATCAGGGGCGCGGCGCTGGGCGCAGAGTGGCGGCGCAAGGCGGCTCGGG 772  
DB 2117 CCTGGATTGAGAAGACCTTGGAGGAGGCGGAGGCGGGGTACGTGGAGACCTCTTGG 2176  
QY 773 GGGACAAGAGCCACTT 788  
DB 2177 GCCCGCGGCTACGT 2192

RESULT 8  
US-08-254-359A-7  
Sequence 7, Application US/08254359A  
Patent No. 5614402  
GENERAL INFORMATION:  
APPLICANT: DAHLBERG, JAMES E.  
APPLICANT: LYAMICHEV, VICTOR I.  
APPLICANT: BROW, MARY ANN D.  
TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO

STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/254,359A  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/073,384  
FILING DATE: 06-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/986,330  
FILING DATE: 07-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: CARROLL, PETER G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: FORS-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2502 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-254-359A-7

Query Match 1.2%; Score 44.8; DB 1; Length 2502;  
Best Local Similarity 46.8%; Pred. No. 0.06;  
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;  
QY 413 TGATGCTGAGAGCAATAAGTCGCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACC 472  
DB 1820 TGGTGGCCCTGGACTATAGCCAGATAGACTCCGGGTCTCTGCCACCTCTCCGGGGAGG 1879  
QY 473 AGCGCTGTGTGGAGCCTTCGAGGGGAGCGGAGCAGCATCTCCGGGGGCGCATCACT 532  
DB 1880 AGAAGCTGATCGGTCTTCCAGGAGGGAGGAGCATCCACACCGAGCGGCGCAAGACCATCA 1939  
QY 533 TCAGCAGCAGATTGCTGTCCGACCGCGCCACAGGTCTTCTCCAGGCGCAGCGCGAGG 592  
DB 1940 TGTTCGGCTCCCGGAGGCGGTGGACCCCTGATGGCGGGCGGCGCAAGACCATCA 1999  
QY 593 AGAGCGGATCTGCTCCAGACCTGTCTCTCCGACCGCCACCTGGCCAGCGGCGCATC 652  
DB 2000 ACTTCGGGTC---CTCTACGGCATGTCCGCCACCGCTCTCCAGGAGCTTGCATCC 2056  
QY 653 TGGAGACCGAGTGGTTCACGCGCTCCGCGCAAGTGGAGGCGCCACTTACACCGCGG 712  
DB 2057 CCTACGAGGCGGTGGCTTCTTATGAGCGCTACTTCCAGAGCTTCCCAAGGTGCGGG 2116  
QY 713 GCCCAATCAGGGGCGCGGCGCTGGGCGCAGAGTGGCGGCGCAAGGCGGCTCGGG 772  
DB 2117 CCTGGATTGAGAAGACCTTGGAGGAGGCGGAGGCGGGGTACGTGGAGACCTCTTGG 2176  
QY 773 GGGACAAGAGCCACTT 788  
DB 2177 GCCCGCGGCTACGT 2192  
RESULT 9  
US-08-483-043-7  
Sequence 7, Application US/08483043  
Patent No. 5691142  
GENERAL INFORMATION:  
APPLICANT: DAHLBERG, JAMES E.









```

; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-599-491-7

Query Match          1.2%; Score 44.8; DB 2; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.06;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

QY 413 TGATGCTGCAGAGCAATAAGTCGCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACC 472
Db 1820 TGTGCGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCTCTGGCCACCTCTCCGGGGAGC 1879

QY 473 AGCGCGTGTGTGGAGCCTTCTGGAGGGGAGCCAGCATCTCCCGGGGGCCATCACCT 532
Db 1880 AGAACCTGATCCGGGTCTTCCAGAGGGGAGGAGATCCACACCCAGCCAGCTGGA 1939

QY 533 TCAGCACCGGATTCGCTGCCAGCCGCCACAGGCTCTTCTCCAGGCCACCGCGGAGG 592
Db 1940 TGTTCGGGTCTCCCGGGAGCGGTGACCCCTGATGCGCGGGCGGCAAGACCATCA 1999

QY 593 AGAGCGCATCTGCTCCAGAGCTCTCTCTCCGACCCACCTTCCAGAGGCGGCGGCTC 652
Db 2000 ACTTCGGGTC---CTCTACGGCATCTCCGCGGCGGCGGCGGCGGCGGCGG 2056

QY 653 TGAGAGCAGAGTGTTCACAGGCTCCGCGGCAAGTGGAGGCGGCGGCTTACACCGCGG 712
Db 2057 CTACGAGGAGGCGGTGGGCTTCAATTGAGCGGTACTTCCAGAGCTTCCCAAGGTGCGG 2116

QY 713 GCCCAATCAGGGGCGCGGCGCTTCCGCGCACAGCTGCGGCGGCGGCGGCGGCGG 772
Db 2117 CTGATTGAGAGACCTTGGAGGAGGCGGAGGCGGCGGCGGCTACGTGGAGACCTCTTCG 2176

QY 773 GGGACAAGAGCCACTT 788
Db 2177 GCGCGCGGCGCTACGT 2192

RESULT 15
US-08-756-386-7
; Sequence 7, Application US/08756386
; Patent No. 5985557
; GENERAL INFORMATION:
; APPLICANT: Prudent, James R.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichiev, Victor I.
; TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Cartoll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,386
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02564
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-756-386-7

Query Match          1.2%; Score 44.8; DB 2; Length 2502;
Best Local Similarity 46.8%; Pred. No. 0.06;
Matches 176; Conservative 0; Mismatches 197; Indels 3; Gaps 1;

QY 413 TGATGCTGCAGAGCAATAAGTCGCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACC 472
Db 1820 TGTGCGCCCTGGACTATAGCCAGATAGAGCTCCGGGTCTCTGGCCACCTCTCCGGGGAGC 1879

QY 473 AGCGCGTGTGTGGAGCCTTCTGGAGGGGAGCCAGCATCTCCCGGGGGCCATCACCT 532
Db 1880 AGAACCTGATCCGGGTCTTCCAGAGGGGAGGAGATCCACACCCAGCCAGCTGGA 1939

QY 533 TCAGCACCGGATTCGCTGCCAGCCGCCACAGGCTCTTCTCCAGGCCACCGCGGAGG 592
Db 1940 TGTTCGGGTCTCCCGGGAGCGGTGACCCCTGATGCGCGGGCGGCAAGACCATCA 1999

QY 593 AGAGCGCATCTGCTCCAGAGCTCTCTCTCCGACCCACCTTCCAGAGGCGGCGGCTC 652
Db 2000 ACTTCGGGTC---CTCTACGGCATCTCCGCGGCGGCGGCGGCGGCGGCGG 2056

QY 653 TGAGAGCAGAGTGTTCACAGGCTCCGCGGCAAGTGGAGGCGGCGGCTTACACCGCGG 712
Db 2057 CTACGAGGAGGCGGTGGGCTTCAATTGAGCGGTACTTCCAGAGCTTCCCAAGGTGCGG 2116

QY 713 GCCCAATCAGGGGCGCGGCGCTTCCGCGCACAGCTGCGGCGGCGGCGGCGGCGG 772
Db 2117 CTGATTGAGAGACCTTGGAGGAGGCGGAGGCGGCGGCGGCTACGTGGAGACCTCTTCG 2176

QY 773 GGGACAAGAGCCACTT 788
Db 2177 GCGCGCGGCGCTACGT 2192

```

Search completed: February 13, 2003, 00:15:06  
Job time : 302 secs





QY 241 GAGTGGCCCTTTACCTACACGGGGGACTCCACCAGCTGAAGGAGCCAACTGCTCC 300  
DB 241 GAGTGGCCCTTTACCTACACGGGGGACTCCACCAGCTGAAGGAGCCAACTGCTCC 300  
QY 301 GGCCGCTAGAGTTGGGGGCTTGGCGGGAAGTGCCAGCCCTTGCCAGCGGCGACCC 360  
DB 301 GGCCGCTAGAGTTGGGGGCTTGGCGGGAAGTGCCAGCCCTTGCCAGCGGCGACCC 360  
QY 361 TCCCTGCACCGGCGCTGGACACACTGCACACGCCACCCCAACTTCCTCAACGTGATGCTG 420  
DB 361 TCCCTGCACCGGCGCTGGACACACTGCACACGCCACCCCAACTTCCTCAACGTGATGCTG 420  
QY 421 CAGACAAATAAGTCGGGGAGCAGAACTTGCAGGAGCACTGGATTGGTACAGGGCGTG 480  
DB 421 CAGACAAATAAGTCGGGGAGCAGAACTTGCAGGAGCACTGGATTGGTACAGGGCGTG 480  
QY 481 GTGTGAGGCTTCTGAGGGCGAGCCACGATCTCCGGGGCGCCATCACTTTCAGCAC 540  
DB 481 GTGTGAGGCTTCTGAGGGCGAGCCACGATCTCCGGGGCGCCATCACTTTCAGCAC 540  
QY 541 GATTTCGCTGTCGCGACGGGCCCCACAGGCTTCCTCCAGGCCACGGCGAGGAGCCGC 600  
DB 541 GATTTCGCTGTCGCGACGGGCCCCACAGGCTTCCTCCAGGCCACGGCGAGGAGCCGC 600  
QY 601 ATCCTGCTCCAAGACCTGTCTCTCCGCGACCCCACTGGCCAAACGCCACTCTGGAGAC 660  
DB 601 ATCCTGCTCCAAGACCTGTCTCTCCGCGACCCCACTGGCCAAACGCCACTCTGGAGAC 660  
QY 661 GAGTGTTCACGGCTCCGGGCGCAAGTGGAGGCCCACTTACACGGCGCGGCCCAAT 720  
DB 661 GAGTGTTCACGGCTCCGGGCGCAAGTGGAGGCCCACTTACACGGCGCGGCCCAAT 720  
QY 721 CAGGGCCCCGGGCGCTGGGCCACAGCTGCGGCGCAAGGCGGCTCGCGGGGACAG 780  
DB 721 CAGGGCCCCGGGCGCTGGGCCACAGCTGCGGCGCAAGGCGGCTCGCGGGGACAG 780  
QY 781 AGCCACTTCAAGTGTCTCGGCTTATCTGAGTGGGAGAGGGAGTTACAGCCCGG 840  
DB 781 AGCCACTTCAAGTGTCTCGGCTTATCTGAGTGGGAGAGGGAGTTACAGCCCGG 840  
QY 841 TGGCTGGTTACTCTTTCCTCTGCGATCTACGGGTTCAGCCCTAACCTGTCGGGAATTC 900  
DB 841 TGGCTGGTTACTCTTTCCTCTGCGATCTACGGGTTCAGCCCTAACCTGTCGGGAATTC 900  
QY 901 AGGGGTGTCATGAAGTTGACATAAATCTCAGAAAGTGGACATTCACCAATGCTCAAGT 960  
DB 901 AGGGGTGTCATGAAGTTGACATAAATCTCAGAAAGTGGACATTCACCAATGCTCAAGT 960  
QY 961 GATGGCTGTTTTCAGGAACCTCAATAATGCCACCTCAACAATTCAGAGTGTATGCCAAT 1020  
DB 961 GATGGCTGTTTTCAGGAACCTCAATAATGCCACCTCAACAATTCAGAGTGTATGCCAAT 1020  
QY 1021 AAAGGCTTAGGATTCGTTCTTGAGGCTATGATGATTTGCAAGAGGATTCATCAT 1080  
DB 1021 AAAGGCTTAGGATTCGTTCTTGAGGCTATGATGATTTGCAAGAGGATTCATCAT 1080  
QY 1081 CTTGGAGTCTTACCAGTGAACAACTTCGGAAGGGGTTCGGATCAGCATATTTTCAGGA 1140  
DB 1081 CTTGGAGTCTTACCAGTGAACAACTTCGGAAGGGGTTCGGATCAGCATATTTTCAGGA 1140  
QY 1141 AGTACAAAAGATGTGTGAGGAAGCCTATGCTGCCTACCTTGCAGGAGGGCTGCCCC 1200  
DB 1141 AGTACAAAAGATGTGTGAGGAAGCCTATGCTGCCTACCTTGCAGGAGGGCTGCCCC 1200  
QY 1201 TTCTGTGCTGATGACAGCCCATGCTTGTCCAGGAGATAAGTATTTAGGACTTGCCTATC 1260  
DB 1201 TTCTGTGCTGATGACAGCCCATGCTTGTCCAGGAGATAAGTATTTAGGACTTGCCTATC 1260  
QY 1261 ATCTCTCTCCAAGGCTGTGTATGCTGCTGACCTTCTGTAGCATGCTGTGTCTTACCAC 1320  
DB 1261 ATCTCTCTCCAAGGCTGTGTATGCTGCTGACCTTCTGTAGCATGCTGTGTCTTACCAC 1320  
QY 1321 TTTCGAAAAGCAAGAGCATCCGGGCATCGGGCCTTATCTCTTTGGAAACGATCCTTTT 1380

DB 1321 TTTCGAAAAGCAAGAGCATCCGGGCATCGGGCCTTATCTCTTGGAAACGATCCTTTT 1380  
QY 1381 GGATCTCTGCTCCTATACCTTCCAGTGTGTTATTTGTACTTTGAGCCAAAGCACATTTCCG 1440  
DB 1381 GGATCTCTGCTCCTATACCTTCCAGTGTGTTATTTGTACTTTGAGCCAAAGCACATTTCCG 1440  
QY 1441 TGTATTCTCTAAGATGGGCTGCTCTCTCGGTTTGTACTGTTTACGGAACGTGCTACT 1500  
DB 1441 TGTATTCTCTAAGATGGGCTGCTCTCTCGGTTTGTACTGTTTACGGAACGTGCTACT 1500  
QY 1501 CTCAAACTTTCACAGGGTTTGAAGGTGTTTCTTTCACGAACGGCTCAACGAATTTCCATAT 1560  
DB 1501 CTCAAACTTTCACAGGGTTTGAAGGTGTTTCTTTCACGAACGGCTCAACGAATTTCCATAT 1560  
QY 1561 ATGACTTGGGGGACGGGTGATGAGGATGCTGGCAGTAATACTCTTGGTAGTGTGTTGGTTT 1620  
DB 1561 ATGACTTGGGGGACGGGTGATGAGGATGCTGGCAGTAATACTCTTGGTAGTGTGTTGGTTT 1620  
QY 1621 CTCATTGGCTGACATCTGCTGCGCAGAAATTTGGAGAAACAGATTTTCACTTATTGGC 1680  
DB 1621 CTCATTGGCTGACATCTGCTGCGCAGAAATTTGGAGAAACAGATTTTCACTTATTGGC 1680  
QY 1681 CAGGGAAAAACATCCGATCACCTCATCTTCAATATGTGCTCATTTGACCGCTGGGACTAC 1740  
DB 1681 CAGGGAAAAACATCCGATCACCTCATCTTCAATATGTGCTCATTTGACCGCTGGGACTAC 1740  
QY 1741 ATGACAGCAGTTGCTGAATTTTATCTCTTGTGGGTGTTTATCTCTGCTATGCAAGT 1800  
DB 1741 ATGACAGCAGTTGCTGAATTTTATCTCTTGTGGGTGTTTATCTCTGCTATGCAAGT 1800  
QY 1801 CGGACAGTCCCATCGSCATTCATGAGCCCGCTATATGGCTGTTGCGATTCACAAATGAG 1860  
DB 1801 CGGACAGTCCCATCGSCATTCATGAGCCCGCTATATGGCTGTTGCGATTCACAAATGAG 1860  
QY 1861 CTCATCATCTCTGCTATATTCATACAATAGATTGTTCTTGCCTCAAGACTTCAGTCT 1920  
DB 1861 CTCATCATCTCTGCTATATTCATACAATAGATTGTTCTTGCCTCAAGACTTCAGTCT 1920  
QY 1921 GATTGGATGTTGATGCTGTATTTTCACATACCTATCTGCTGACAGTCAACATTTGGG 1980  
DB 1921 GATTGGATGTTGATGCTGTATTTTCACATACCTATCTGCTGACAGTCAACATTTGGG 1980  
QY 1981 TTGCTTTTGTATTCAAAAGTTTTCACATTCAAGCAATAACCCAGAGATGATATTCTACA 2040  
DB 1981 TTGCTTTTGTATTCAAAAGTTTTCACATTCAAGCAATAACCCAGAGATGATATTCTACA 2040  
QY 2041 GAAGCATATGAGGATGAGTACATAGACATGGGCGGATCTGGATCCTTACCTGAACAGCATC 2100  
DB 2041 GAAGCATATGAGGATGAGTACATAGACATGGGCGGATCTGGATCCTTACCTGAACAGCATC 2100  
QY 2101 AATTGAGCTGAGTACAGCAGCTTGGATCAGAGGACATTTGGGAGCAGCTGAAAAA 2160  
DB 2101 AATTGAGCTGAGTACAGCAGCTTGGATCAGAGGACATTTGGGAGCAGCTGAAAAA 2160  
QY 2161 CTCTATGCCCACTGGAATATATAAAGAAAGAGATGATCACAACAACCCACCTC 2220  
DB 2161 CTCTATGCCCACTGGAATATATAAAGAAAGAGATGATCACAACAACCCACCTC 2220  
QY 2221 CAGAAAAAGCGGTGCTCGAAGAAGGGCCTAGTCTGTTCCATCATGAGACGATTTACGGAG 2280  
DB 2221 CAGAAAAAGCGGTGCTCGAAGAAGGGCCTAGTCTGTTCCATCATGAGACGATTTACGGAG 2280  
QY 2281 ATCCAGAGACATGACCGGCGAGTGTCTTAAGAGGACAAGGAGGGCGCGGACCATGGC 2340  
DB 2281 ATCCAGAGACATGACCGGCGAGTGTCTTAAGAGGACAAGGAGGGCGCGGACCATGGC 2340  
QY 2341 ACAGCAAAAGCAGTGCCTCATCAGGAAGAACCCCGCAGAGTCTTTCAGGAAACACAGG 2400  
DB 2341 ACAGCAAAAGCAGTGCCTCATCAGGAAGAACCCCGCAGAGTCTTTCAGGAAACACAGG 2400  
QY 2401 AAATCCAAAGGAGGACCTGAAAAAACCCGAGTCTTCTCACTCAAGAAATCCACAGCACT 2460  
DB 2401 AAATCCAAAGGAGGACCTGAAAAAACCCGAGTCTTCTCACTCAAGAAATCCACAGCACT 2460



|||||  
Db 969 GATTCTGTCGCGACCGCCACAGGCTCTCTCCAGGCCACGCGGAGGAGCCGC 1028  
QY 601 ATCCTGCTCCAAGACCTGTCCTCTCCGACACCCACCTGGCCAAACCCACTCTGGAGACC 660  
Db 1029 ATCCTGCTCCAAGACCTGTCCTCTCCGACACCCACCTGGCCAAACCCACTCTGGAGACC 1088  
QY 661 GAGTGTGTCACGGCTCCGGGCGCAAGTGGAGGCCCCACCTTACACCGCGCGGCCCAAT 720  
Db 1089 GAGTGTGTCACGGCTCCGGGCGCAAGTGGAGGCCCCACCTTACACCGCGCGGCCCAAT 1148  
QY 721 CAGGGCCCCGGGCGCTGGGCCACAGCTGGCGGCGCAAGGAGGGCTCGCGGGGACAAAG 780  
Db 1149 CAGGGCCCCGGGCGCTGGGCCACAGCTGGCGGCGCAAGGAGGGCTCGCGGGGACAAAG 1208  
QY 781 AGCCACTTCAAGTGTCTCGGCTTATCTGGAGTGGGAGAAAGGGAGTTACAAAGCCCGGG 840  
Db 1209 AGCCACTTCAAGTGTCTCGGCTTATCTGGAGTGGGAGAAAGGGAGTTACAAAGCCCGGG 1268  
841 TGGCTGGTTACTCTTTCCTCTGCGCATCTACGGGTTGACGCTAACCTGGTCCCGGAATTC 900  
Db 1269 TGGCTGGTTACTCTTTCCTCTGCGCATCTACGGGTTGACGCTAACCTGGTCCCGGAATTC 1328  
QY 901 AGGGGTGTCATGAAGTGTGACATAAATCTTCAGAAAGTGGACATTTGACCAATGCTCAAGT 960  
Db 1329 AGGGGTGTCATGAAGTGTGACATAAATCTTCAGAAAGTGGACATTTGACCAATGCTCAAGT 1388  
QY 961 GATGGCTGGTTTCAGGAACCTATAAATGCCACCTCAACAAATTCAGAGTGTATGCCAAT 1020  
Db 1389 GATGGCTGGTTTCAGGAACCTATAAATGCCACCTCAACAAATTCAGAGTGTATGCCAAT 1448  
QY 1021 AAAGGCTTAGGATTCCTCTTGAGGCTATGAGTGCATTTGCAAGCAGGATTCATCAT 1080  
Db 1449 AAAGGCTTAGGATTCCTCTTGAGGCTATGAGTGCATTTGCAAGCAGGATTCATCAT 1508  
QY 1081 CTTGGAGTCTTACCAGTGAACAACTTTCGGAGAAAGGGTCCGGATCAGACATATTTCAAG 1140  
Db 1509 CTTGGAGTCTTACCAGTGAACAACTTTCGGAGAAAGGGTCCGGATCAGACATATTTCAAG 1568  
QY 1141 AGTACAAAAGATGTGTCAGAAAGGCTATGCTGCCTACCTTGCAGGAGGGCTGCCCC 1200  
Db 1569 AGTACAAAAGATGTGTCAGAAAGGCTATGCTGCCTACCTTGCAGGAGGGCTGCCCC 1628  
QY 1201 TTCTGTGCTGATGACAGCCCATGCTGCTCCAGGAAGATAGTATTACGACTTGCCATC 1260  
Db 1629 TTCTGTGCTGATGACAGCCCATGCTGCTCCAGGAAGATAGTATTACGACTTGCCATC 1688  
QY 1261 ATCTCCTTCCAAGGCTGTGTATGCTGCTGACCTTTCGTTAGCATGCTGTGTCTACCCAC 1320  
Db 1689 ATCTCCTTCCAAGGCTGTGTATGCTGCTGACCTTTCGTTAGCATGCTGTGTCTACCCAC 1748  
QY 1321 TTTCGCAAGCAAGACATCCGGGATCCGGGCTATCTCTGTTGAAACGATCCTTTT 1380  
Db 1749 TTTCGCAAGCAAGACATCCGGGATCCGGGCTATCTCTGTTGAAACGATCCTTTT 1808  
QY 1381 GGATCTCTCTCTTACATTTCCAGTTGTTATTTGTACTTTTGGCCAAAGCAGCATTTCCG 1440  
Db 1809 GGATCTCTCTCTTACATTTCCAGTTGTTATTTGTACTTTTGGCCAAAGCAGCATTTCCG 1868  
QY 1441 TGTATTCTCTTAAGATGGGCTGCTTCTCGGTTTTCGTTTGTACTTGTACGGAACTGTCACT 1500  
Db 1869 TGTATTCTCTTAAGATGGGCTGCTTCTCGGTTTTCGTTTGTACTTGTACGGAACTGTCACT 1928  
QY 1501 CTCAAACTTTCAGGGTTTTGAAGTGTCTTTCACGAACGGCTCAACGAATTCATAT 1560  
Db 1929 CTCAAACTTTCAGGGTTTTGAAGTGTCTTTCACGAACGGCTCAACGAATTCATAT 1988  
QY 1561 ATGACTGGCGGACGGTCTATGAGGTGCTGGCAGTAATACTCTTGGTACTGTTTTGGTT 1620  
Db 1989 ATGACTGGCGGACGGTCTATGAGGTGCTGGCAGTAATACTCTTGGTACTGTTTTGGTT 2048  
QY 1621 CTCAATGGCTGGACTTCATCTGTGCCAGAAATTTGGAGAAACAGATTTCACTTATTGGC 1680  
|||||

Db 2049 CTCAATGGCTGGACTTCATCTGTGTGCCAGAATTTGGAGAAACAGATTTCACTTATTGGC 2108  
QY 1681 CAGGGGAAACATCCGATCACCTCATCTCAATATCTGCTCATCTGCTCATCGCGTGGGACTAC 1740  
Db 2109 CAGGGGAAACATCCGATCACCTCATCTCAATATCTGCTCATCTGCTCATCGCGTGGGACTAC 2168  
QY 1741 ATTGACAGCAGTTG----- 1753  
Db 2169 ATGACAGCAGTTGGTATGTGGTCACTTGTTCGTATGATGGTCTTACCATTTTTTCAGTAG 2228  
QY 1754 ----- 1753  
Db 2229 CATTACACTGACTCTTTTAAAAGCTTGGCCAGACTCTTTAGTAAAGCCTAAACTAATCT 2288  
QY 1754 -----CTGAAATTTTATTCCTTGTGGGGTGTATCTCTGCTATGCGAGTGGCGACAG 1807  
Db 2289 CACCTACTGAATTTTATTCCTTGTGGGGTGTATCTCTGCTATGCGAGTGGCGACAG 2348  
QY 1808 TCCCATCGGCATTCATGAGCCCGCTATATGGCTGTTCAGTTTCCAGTTTCAATGAGCTCATCA 1867  
Db 2349 TCCCATCGGCATTCATGAGCCCGCTATATGGCTGTTCAGTTTCAATGAGCTCATCA 2408  
QY 1868 TCTCTGTATATTCATACAAATAGATTTGTCTTGGCTCAAGACTTTCAGTCTGATTTGA 1927  
Db 2409 TCTCTGTATATTCATACAAATAGATTTGTCTTGGCTCAAGACTTTCAGTCTGATTTGA 2468  
QY 1928 TGTGATGCTGTATTTTGCACATACTCATTTGACGTGTGACAGTCCACCATTTGGGTTGCTTT 1987  
Db 2469 TGTGATGCTGTATTTTGCACATACTCATTTGACGTGTGACAGTCCACCATTTGGGTTGCTTT 2528  
QY 1988 TGATTTCCAAAGTTTTCACATTTCAAGCAATAACCCAGAGATGATATTGCTACAGAAAGCAT 2047  
Db 2529 TGATTTCCAAAGTTTTCACATTTCAAGCAATAACCCAGAGATGATATTGCTACAGAAAGCAT 2588  
QY 2048 ATGAGGATGAGCTAGACATAGGCGCGATCTGGATCCTACCTGAACAGCAGTATCAATTCAG 2107  
Db 2589 ATGAGGATGAGCTAGACATAGGCGCGATCTGGATCCTACCTGAACAGCAGTATCAATTCAG 2648  
QY 2108 CCGTGAAGTGAACAGCTTGGATCCAGAGGACATTTGGGACGAGCTGAAACAACTCTATG 2167  
Db 2649 CCGTGAAGTGAACAGCTTGGATCCAGAGGACATTTGGGACGAGCTGAAACAACTCTATG 2708  
QY 2168 CCCAACTGGAAATATATAAAGAAAGAGATGATCACAAACAAACCCACCTCCAGAAAA 2227  
Db 2709 CCCAACTGGAAATATATAAAGAAAGAGATGATCACAAACAAACCCACCTCCAGAAAA 2768  
QY 2228 AGCGTGTCTGAAGAGGGCCTAGGTCTTCATCATGAGAGCGCATTTAGGAGATCCGAG 2287  
Db 2769 AGCGTGTCTGAAGAGGGCCTAGGTCTTCATCATGAGAGCGCATTTAGGAGATCCGAG 2828  
QY 2288 AGACAGTCAAGCGGAGTCTCTAAAGAGGACAAGGCGGCCGCCACCATGGCAGAGCA 2347  
Db 2829 AGACAGTCAAGCGGAGTCTCTAAAGAGGACAAGGCGGCCGCCACCATGGCAGAGCA 2888  
QY 2348 AAGGCACTGCCCTCATCAGGAAGAACCCGCCAGAGTCTTTCAGGGAACACAGGGAATCCA 2407  
Db 2889 AAGGCACTGCCCTCATCAGGAAGAACCCGCCAGAGTCTTTCAGGGAACACAGGGAATCCA 2948  
QY 2408 AGGAGGAGACCTGAAACACCGAGTCTTCTACTCAAGAAATCCACAGACATTTATGACC 2467  
Db 2949 AGGAGGAGACCTGAAACACCGAGTCTTCTACTCAAGAAATCCACAGACATTTATGACC 3008  
QY 2468 AGCTGAGAGCAACAGGAAGTCCAGTAGCTTACCACAGAAAGCAAGAGGAGGAGA 2527  
Db 3009 AGCTGAGAGCAACAGGAAGTCCAGTAGCTTACCACAGAAAGCAAGAGGAGGAGA 3068  
QY 2528 CAACAGAAATTTCCACACTTGAATCCCTCTCGGGTAAAAAACTAAACAAAAAATCAAAAG 2587  
Db 3069 CAACAGAAATTTCCACACTTGAATCCCTCTCGGGTAAAAAACTAAACAAAAAATCAAAAG 3128  
QY 2588 AAGACAGCAGGCTGAGTCCAGGAGTCCGTCGCTTGGTGTGCAAGTACAGCAAGCGCTC 2647  
Db 3129 AAGACAGCAGGCTGAGTCCAGGAGTCCGTCGCTTGGTGTGCAAGTACAGCAAGCGCTC 3188  
|||||





```
Db 721 CAGGGCCCGGGGCTGGGCCACAGCTGGCGGCGCAAGACGGGCTCGCGGGGACAAAG 780
QY 781 AGCCACTCAAGTGNCTCCGGCTTATCTGAGTGGAGAAAGGGAGTTACAAGCCCGG 840
Db 781 AGCCACTCAAGTGGTCTCCGGCTTATCTGAGTGGAGAAAGGGAGTTACAAGCCCGG 840
QY 841 TGGCTGTTTACTCTTTCCTCTCCATCTACGGGTTTCAGCCTAACCTGTGTCGGGAATTC 900
Db 841 TGGCTGTTTACTCTTTCCTCTCCATCTACGGGTTTCAGCCTAACCTGTGTCGGGAATTC 900
QY 901 AGGGGTGCATGAAGTTGACATAAATCTCAGAAAGTGGATGACCAATGCTCAAGT 960
Db 901 AGGGGTGCATGAAGTTGACATAAATCTCAGAAAGTGGATGACCAATGCTCAAGT 960
QY 961 GATGGCTGTTTTCAGGAATCTAATAATCCCACTCAACAATTCAGAGTGTATGCCAAT 1020
Db 961 GATGGCTGTTTTCAGGAATCTAATAATCCCACTCAACAATTCAGAGTGTATGCCAAT 1020
QY 1021 AAAGGCTTAGGATTCCTTCTTGAGGCTATGAGTGCATTTGCAAGCAGGATTCATCAT 1080
Db 1021 AAAGGCTTAGGATTCCTTCTTGAGGCTATGAGTGCATTTGCAAGCAGGATTCATCAT 1080
QY 1081 CTTGGAGTCTTACAGTGAACAACCTTTCGGAAGAGGGTCCGGATCAGCATATTTACAGA 1140
Db 1081 CTTGGAGTCTTACAGTGAACAACCTTTCGGAAGAGGGTCCGGATCAGCATATTTACAGA 1140
QY 1141 AGTACAAAAGATGTGTACAGAAAGCCTATGCTGCCTACCTTGCAGGAGGGCTGCCCC 1200
Db 1141 AGTACAAAAGATGTGTACAGAAAGCCTATGCTGCCTACCTTGCAGGAGGGCTGCCCC 1200
QY 1201 TTCTGTGCTGATGACAGCCCATGCTTCGTCAGGAAGATAAGTATTACGACTTGCCATC 1260
Db 1201 TTCTGTGCTGATGACAGCCCATGCTTCGTCAGGAAGATAAGTATTACGACTTGCCATC 1260
QY 1261 ATCTCCTCCAAAGGCTGTGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
Db 1261 ATCTCCTCCAAAGGCTGTGTATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
QY 1321 TTTCGCAAGCAAGAGATCCGGGATCGGGCTATTCCTGTTGGAACGATCCTTTT 1380
Db 1321 TTTCGCAAGCAAGAGATCCGGGATCGGGCTATTCCTGTTGGAACGATCCTTTT 1380
QY 1381 GGATCTCTGCTCTATACCTTTCAGTGTGTTATTTGTTGTTGTTGTTGTTGTTGTTGTT 1440
Db 1381 GGATCTCTGCTCTATACCTTTCAGTGTGTTATTTGTTGTTGTTGTTGTTGTTGTTGTT 1440
QY 1441 TGTATTTCTCCTTAAGATGGGCTCGTCTCTCGGTTTGTGCTGTTGTTGTTGTTGTTGTT 1500
Db 1441 TGTATTTCTCCTTAAGATGGGCTCGTCTCTCGGTTTGTGCTGTTGTTGTTGTTGTTGTT 1500
QY 1501 CTCAACTTCACAGGTTTGAAGTGTTCCTTTCACGACGGCTCAACGAATTCATAT 1560
Db 1501 CTCAACTTCACAGGTTTGAAGTGTTCCTTTCACGACGGCTCAACGAATTCATAT 1560
QY 1561 ATGACTTGGCGGACGGTTCATAGGATGCTGGCAGTAATCTTGTGTTGTTGTTGTTGTT 1620
Db 1561 ATGACTTGGCGGACGGTTCATAGGATGCTGGCAGTAATCTTGTGTTGTTGTTGTTGTT 1620
QY 1621 CTCATTGGCTGACATCTCTGTGTCAGAAATTTGGAGAAACAGATTTCACTTATGGC 1680
Db 1621 CTCATTGGCTGACATCTCTGTGTCAGAAATTTGGAGAAACAGATTTCACTTATGGC 1680
QY 1681 CAGGGAAACATCCGATCCTCATCTTCAATATGTGCTCATTTGACCGCTGGGACTAC 1740
Db 1681 CAGGGAAACATCCGATCCTCATCTTCAATATGTGCTCATTTGACCGCTGGGACTAC 1740
QY 1741 ATGACAGCAGTTGGTATGTGGTCACTTGTTCGTATGATGT 1782
Db 1741 ATGACAGCAGTTGGTATGTGGTCACTTGTTCGTATGATGT 1782
```

RESULT 4  
US-09-764-898-82

```
; Sequence 82, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 2549
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-898-82

Query Match          1.4%; Score 51.2; DB 10; Length 2549;
Best Local Similarity 48.0%; Pred. No. 0.00064;
Matches 146; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 500 GCGAGCCGACGATCTCCGGGGGCGCATCACCTTCAGCACCGATTTCGTGTCCGACCGG 559
Db 1616 GCGAGTGTGCGCGGCTGGGTGCCCGGGCCCTCCGGCCGCGATTGCCAGCTCCCGGAGA 1675

QY 560 CCCACACAGTCTTCTCCAGGCCAGCGCGGAGGAGCGCATCTCTCCAAAGACCTGT 619
Db 1676 GCGCGGCGGACGCTGCCCGGAGCGCGCGCTGTGTGGAGCGCGCGGAGAGGACG 1735

QY 620 CTTCTCTCGACCCACCTGGCCAAACGCACTCTTGGAGACGAGTGTTCACGCGCTCC 679
Db 1736 GCATGTGCTGTGGCCCTTGGCGGGCCCTGCCTCACCTTCGACGACTGCTGCTGCCGCC 1795

QY 680 GCGCAAGTGGAGCGCCCACTTACACCGCGCGGCCCAATCAGGGGCCCGCGGCGCTGG 739
Db 1796 AGGGCGCGGCTGGGGCGCCCAATGCCGACCGTGTCCGCGCGCGGGGTCCCAT 1855

QY 740 GCCACAGTGGCGGCGCAAGGACGGGCTCGGGGGGAGCAAGGCCACTTCAAGTGGTCTC 799
Db 1856 GCGCGACATCGCAGAGAGAGCAATTCCTTCTGGGACACAAAGCCCCCTGTGTGGGGA 1915

QY 800 CGCC 803
Db 1916 AGCC 1919
```

RESULT 5  
US-09-999-183-4  
; Sequence 4, Application US/09999183  
; Patent No. US20020147169A1  
; GENERAL INFORMATION:  
; APPLICANT: MITROPHANOUS, et al  
; TITLE OF INVENTION: In Vivo Selection Method  
; FILE REFERENCE: 674523-2009  
; CURRENT APPLICATION NUMBER: US/09/999,183  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: PCT/GB00/02136  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 9912965.2  
; PRIOR FILING DATE: 1999-06-03  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: SeqWin99  
; SEQ ID NO 4  
; LENGTH: 2571  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Codon optimised env sequence  
US-09-999-183-4

Query Match 1.3%; Score 48; DB 10; Length 2571;  
Best Local Similarity 45.2%; Pred. No. 0.0053;  
Matches 177; Conservative 0; Mismatches 215; Indels 0; Gaps 0;





```

; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-402J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42999 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-125-767-17

Query Match 1.3%; Score 47; DB 9; Length 42999;
Best Local Similarity 48.0%; Pred. No. 0.081;
Matches 134; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 2561 GTAAAAAAGTACACAAAAAAGTAAAGACAGACGAGGCTGAGTCCACGGAGTCGGTGC 2620
DB 40517 GCAAGCAAGAAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGCAAGCA 40458

QY 2621 CGTTGGTGTGCAAGTACGCAAGCGCTCACACCTCAGCTCAGAGAGAAAGTGGCACC 2680
DB 40457 AGCAAGAAAGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 40398

QY 2681 CAGCAATCGATGTTACAGAGTCTCTCAGTGTCTATAGCAAGCGCAAGGAGAGACTC 2740
DB 40397 AGCAAGCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 40338

QY 2741 TTGGATTAGCTGGAAAGCCAAAGCAGAGGTGTGGAAGAGCAGTAAATCCAGAAAC 2800
DB 40337 AAAGAAAGAAATCGAGAAACAAAGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 40278

QY 2801 CTTTGGCAAAAGTAAAGAGACAAACAGAAATCACTCAA 2839
40277 AGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGCAAGGAAA 40239

RESULT 8
US-10-044-090-309
; Sequence 309, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 309
; LENGTH: 2724
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 300437.18
; NAME/KEY: unsure
; LOCATION: 2060
; OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-309

```

```

Query Match 1.3%; Score 46; DB 12; Length 2724;
Best Local Similarity 52.6%; Pred. No. 0.021;
Matches 100; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 547 CTGTCCGACACCGCCCGCCAGAGTCTTCTCCAGGACACGCGGAGAGAGCCGATCTCTG 606
DB 362 CAGGCGCGCGCGCGCCCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 421

QY 607 CTCCAAGACCTGTCTCTCCGACCCCGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 666
DB 422 GACTTGTCTCCCGGTGCTCGCCCGCCAGGAGTCCGCGCGGCTGCTCGGCTCTGCTGC 481

QY 667 TTCCACGGGCTCGCGGCAAGTGGAGGCGCCACTTACACCGCGCGGCGCCCAATCAGGG 726
DB 482 CGCGGCTGTGCGGCTGCTGCGGCGCCCGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 541

QY 727 CCCCAGGGCC 736
DB 542 CACAGCGGCC 551

RESULT 9
US-09-476-242-22
; Sequence 22, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; CURRENT FILING DATE: 1999-12-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 2298
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Val120-Ile201;
; OTHER INFORMATION: Ile424-Ala433
US-09-476-242-22

Query Match 1.2%; Score 45.6; DB 10; Length 2298;
Best Local Similarity 44.0%; Pred. No. 0.024;
Matches 192; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 359 CCTCCTTGACCGCGCGCTGGACACACTCACACAGCCCGCCACCACTTCTCTCAACGTGATGC 418
DB 1601 CCTGTGAGTGGGAGCGCGAGATCGACACTACCAACCTGATCTACACCTGATCG 1660

QY 419 TGCAGAGCAATTAAGTCGGGGAGCAGAACTTGCAGAGCAGCTTGGATTGGTACCAGGGCC 478
DB 1661 AGGAGAGCCAGAAACAGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1720

QY 479 TGGTGTGAGCCTTCTGGAGGCGGAGCCAGCATCTCCGCGGCGGCCATCACCCTTCAGCA 538
DB 1721 CCAGCCTGTGGAACCTGGTTCGACATCAGCAAGTGGCTGTGGTACATCAAGATCTTCATCA 1780

QY 539 CCGATTTCGCTGTCCGCACCGCGCCCGCCAGAGTCTTCTCTCAGGCCACGCGCGAGGAGGCC 598
DB 1781 TGATCGTGGCGGCTGTGGGCTTGGCGCATCGTGTTCACCTGTGAGCATCGTGAACC 1840

QY 599 GCATCCTGCTCAAGACCTTGTCTCTCGCACCCCGCCAGCCAGCCAGCCAGCCAGCCAG 658
DB 1841 GCGTGCAGGCGCTACAGCCCGCTGAGCTTCAGACCGGCTTCCCGCGCCCGCCCGCGGCC 1900

QY 659 CCGAGTGTTCACGGCCCTCGCGCCCAAGTGGAGGCGGCCACTTACACCGCGCGGCCCA 718
DB 1901 CCGACCGCGCGGAGGAGTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1960

QY 719 ATCAGGGGCGCGCGGCTTGGGCCACAGCTGGCGCGCAAGGAGCGGGCTCGCGCGGGACA 778

```

Db 1961 CCTGGTGCACGGCCTGCTGGCCCTGATCTGGGACGACCTTGGCGACGCTGTGCCTGTTC A 2020  
QY 779 AGAGCCACTTCAAGTG 794  
Db 2021 GCTACCACCGCCTGCG 2036

RESULT 10  
US-09-476-242-23  
; Sequence 23, Application US/09476242  
; Patent No. US20020146683A1  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: HARTOG, Karin  
; APPLICANT: MARTIN, Eric  
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES  
; FILE REFERENCE: 1605.002  
; CURRENT APPLICATION NUMBER: US/09/476,242  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 2298  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: Vall20-Ile201B; Ile424-Ala433  
US-09-476-242-23

Query Match 1.2%; Score 45.6; DB 10; Length 2298;  
Best Local Similarity 44.0%; Pred. No. 0.024;  
Matches 192; Conservative 0; Mismatches 244; Indels 0; Gaps 0;  
QY 359 CTTCTTGCACCGGGCGCTGGACACACTTGACACGCCACCAACTTCCCTCAAGTGATGC 418  
Db 1601 CTTGGATGGAGTGGGCGCGAGATCGACAACCTACCAACCTGTATCTACACCTGTATCG 1660  
QY 419 TGCAGAGCAATAGTCGCGGGAGCAGAACTTGACAGCAGCAGCTGGATTGGTACGAGCGC 478  
Db 1661 AGGAGAGCAGAACCAAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1720  
QY 479 TGGTGTGGAGCCTTCTGGAGGGGAGCGACGATCTCCCGGGCGGCATCACCTTCAGCA 538  
Db 1721 CCAGCCTGTGGAAGTGTTCGACATCAGCAAGTGGCTGTGGTATCATCAAGATCTTCATCA 1780  
QY 539 CCGATTGCTGTCCGACCGGCGCCACAGGTCTTCTCCAGGCCACCGCGGAGGAGCGC 598  
Db 1781 TGATCTGGGCGGCGCTGGTGGGCGCTGCGCATCTGTGTTACCGTGTGAGCATCGTGAACC 1840  
QY 599 GCATCTGTCTCAAGACCTGTCTCTCCGACACCTGGCCCAACGCACTCTCGAGA 658  
Db 1841 GCGTGGCCAGGGCTACAGCCCTTCCAGCCCTTCCAGCCCTTCCCGCCCGCGCGGC 1900  
QY 659 CCGAGTGGTTCACGCGCTCCGCGCAAGTGGAGGCCCACTTACACCGCGCGGCCCA 718  
Db 1901 CCGACCGCCCGAGGCGCATCGAGGAGGAGGCGCGCGACCGCGACCGAGCAGCC 1960  
QY 719 ATCAGGGGCGCGGGCGCTGGGCCACAGCTGGCGGCGCAAGGACGGGTGCGGCGGGGACA 778  
Db 1961 CCTGGTGCACGGCCTGCTGGCCCTGATCTGGGACGACCTTGGCGAGCCTGTGCTGTTC A 2020  
QY 779 AGAGCCACTTCAAGTG 794  
Db 2021 GCTACCACCGCCTGCG 2036

RESULT 11  
US-09-476-242-24  
; Sequence 24, Application US/09476242  
; Patent No. US20020146683A1  
; GENERAL INFORMATION:

; APPLICANT: BARNETT, Susan  
; APPLICANT: HARTOG, Karin  
; APPLICANT: MARTIN, Eric  
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES  
; FILE REFERENCE: 1605.002  
; CURRENT APPLICATION NUMBER: US/09/476,242  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 2298  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Vall20-Thr202;  
; OTHER INFORMATION: Ile424-Ala433  
US-09-476-242-24

Query Match 1.2%; Score 45.6; DB 10; Length 2298;  
Best Local Similarity 44.0%; Pred. No. 0.024;  
Matches 192; Conservative 0; Mismatches 244; Indels 0; Gaps 0;  
QY 359 CTTCTTGCACCGGGCGCTGGACACACTTGACACGCCACCAACTTCCCTCAAGTGATGC 418  
Db 1601 CTTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGTATCTACACCTGTATCG 1660  
QY 419 TGCAGAGCAATAGTCGCGGGAGCAGAACTTGACAGCAGCAGCTGGATTGGTACGAGCGC 478  
Db 1661 AGGAGAGCAGAACCAAGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1720  
QY 479 TGGTGTGGAGCCTTCTGGAGGGGAGCGACGATCTCCCGGGCGGCATCACCTTCAGCA 538  
Db 1721 CCAGCCTGTGGAAGTGTTCGACATCAGCAAGTGGCTGTGGTATCATCAAGATCTTCATCA 1780  
QY 539 CCGATTGCTGTCCGACCGGCGCCACAGGTCTTCTCCAGGCCACCGCGGAGGAGCGC 598  
Db 1781 TGATCTGGGCGGCGCTGGTGGGCGCTGCGCATCTGTGTTACCGTGTGAGCATCGTGAACC 1840  
QY 599 GCATCTGTCTCAAGACCTGTCTCTCCGACACCTGGCCCAACGCACTCTCGAGA 658  
Db 1841 GCGTGGCCAGGGCTACAGCCCTTCCAGCCCTTCCAGCCCTTCCCGCCCGCGCGGC 1900  
QY 659 CCGAGTGGTTCACGCGCTCCGCGCAAGTGGAGGCCCACTTACACCGCGCGGCCCA 718  
Db 1901 CCGACCGCCCGAGGCGCATCGAGGAGGAGGCGCGCGACCGCGACCGAGCAGCC 1960  
QY 719 ATCAGGGGCGCGGGCGCTGGGCCACAGCTGGCGGCGCAAGGACGGGTGCGGCGGGGACA 778  
Db 1961 CCTGGTGCACGGCCTGCTGGCCCTGATCTGGGACGACCTTGGCGAGCCTGTGCTGTTC A 2020  
QY 779 AGAGCCACTTCAAGTG 794  
Db 2021 GCTACCACCGCCTGCG 2036

RESULT 12  
US-09-476-242-3  
; Sequence 3, Application US/09476242  
; Patent No. US20020146683A1  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: HARTOG, Karin  
; APPLICANT: MARTIN, Eric  
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES  
; FILE REFERENCE: 1605.002  
; CURRENT APPLICATION NUMBER: US/09/476,242  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 2310  
; TYPE: DNA  
; ORGANISM: Artificial Sequence

```
;
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Val120-Ala204
US-09-476-242-3

Query Match      1.2%; Score 45.6; DB 10; Length 2310;
Best Local Similarity 44.0%; Pred. No. 0.024;
Matches 192; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 359 CCTCTTGCACCGGGCGCTGGACACACTGACACACGCCACCAACTTCTCAACGTGATGC 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1613 CTGAGTGGAGTGGAGCGCGAGATCGACAACACTACCAACCTGTATCTACACCTGATCG 1672

QY 419 TGCAGAGCAATAAGTCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACAGCGGC 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1673 AGGAGAGCCAGAACACGAGGAGAAAGACGAGCAGAGCTGCTGGAGCTGGACAAGTGGG 1732

QY 479 TGGTGGAGCCCTTCTGGAGGGCGAGCCGACGATCTCCGGGGCGCCATCACTTCCAGCA 538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1733 CCAGCCTGTGGAACCTGTTTCGACATCAGCAAGTGGCTGGTATACATCAAGATCTTCATCA 1792

QY 539 CCGATTGCTGTCGACCGGCCCGGACAGTCTTCTCCAGGCCACCGCGGAGAGGCC 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1793 TGATCGTGGCGGCCCTGGTGGGCCCTGCGCATCGTGTTCACCGTGTGAGCATCGTGAACC 1852

QY 599 GCATCTGCTGCCAAGACCTGTCTCTCCGACACCCACCTGGCCACGCACTCTCTGGAGA 658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1853 GCGTGGCCAGGGCTACAGCCCTCGAGCTTCGAGACCCGCTTCCCGCCCGCCCGCGGCC 1912

QY 659 CCGAGTGGTTCACGCGCTCCGGCGCAAGTGGAGGCCACCTTACACCGCCGCGGCCCA 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1913 CCGACGCCCGGAGGCGCATCGAGGAGGGCGGCGAGCGCGACCGACCGAGCGCC 1972

QY 719 ATCAGGGGCCCGGGCCCTGGGCCACAGCTGGGGGCGCAAGGACGCGGCTCGGGGGGACA 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1973 CCTGTGTCAGCGCCCTGTGGCCCTGATCTGGGACGACCTGGCGACGCTGTGCTGTTC 2032

QY 779 AGAGCCACTTCAAGTG 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2033 GCTACCACCGCCTGCG 2048

RESULT 13
US-09-476-242-21
; Sequence 21, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 2310
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Lys121-Val200;
US-09-476-242-21

Query Match      1.2%; Score 45.6; DB 10; Length 2310;
Best Local Similarity 44.0%; Pred. No. 0.024;
Matches 192; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 359 CCTCTTGCACCGGGCGCTGGACACACTGACACACGCCACCAACTTCTCAACGTGATGC 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1613 CTGAGTGGAGTGGAGCGCGAGATCGACAACACTACCAACCTGTATCTACACCTGATCG 1672

QY 419 TGCAGAGCAATAAGTCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACAGCGGC 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1673 AGGAGAGCCAGAACACGAGGAGAAAGACGAGCAGAGCTGCTGGAGCTGGACAAGTGGG 1732

QY 479 TGGTGGAGCCCTTCTGGAGGGCGAGCCGACGATCTCCGGGGCGCCATCACTTCCAGCA 538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1733 CCAGCCTGTGGAACCTGTTTCGACATCAGCAAGTGGCTGGTATACATCAAGATCTTCATCA 1792

QY 539 CCGATTGCTGTCGACCGGCCCGGACAGTCTTCTCCAGGCCACCGCGGAGAGGCC 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1793 TGATCGTGGCGGCCCTGGTGGGCCCTGCGCATCGTGTTCACCGTGTGAGCATCGTGAACC 1852

QY 599 GCATCTGCTGCCAAGACCTGTCTCTCCGACACCCACCTGGCCACGCACTCTCTGGAGA 658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1853 GCGTGGCCAGGGCTACAGCCCTCGAGCTTCGAGACCCGCTTCCCGCCCGCCCGCGGCC 1912

QY 659 CCGAGTGGTTCACGCGCTCCGGCGCAAGTGGAGGCCACCTTACACCGCCGCGGCCCA 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1913 CCGACGCCCGGAGGCGCATCGAGGAGGGCGGCGAGCGCGACCGACCGAGCGCC 1972

QY 719 ATCAGGGGCCCGGGCCCTGGGCCACAGCTGGGGGCGCAAGGACGCGGCTCGGGGGGACA 778
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1973 CCTGTGTCAGCGCCCTGTGGCCCTGATCTGGGACGACCTGGCGACGCTGTGCTGTTC 2032

QY 779 AGAGCCACTTCAAGTG 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2033 GCTACCACCGCCTGCG 2048

RESULT 14
US-09-476-242-4
; Sequence 4, Application US/09476242
; Patent No. US20020146683A1
; GENERAL INFORMATION:
; APPLICANT: BARNETT, Susan
; APPLICANT: HARTOG, Karin
; APPLICANT: MARTIN, Eric
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES
; FILE REFERENCE: 1605.002
; CURRENT APPLICATION NUMBER: US/09/476,242
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2316
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Val120-Ile201
US-09-476-242-4

Query Match      1.2%; Score 45.6; DB 10; Length 2316;
Best Local Similarity 44.0%; Pred. No. 0.024;
Matches 192; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

QY 359 CCTCTTGCACCGGGCGCTGGACACACTGACACACGCCACCAACTTCTCAACGTGATGC 418
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1619 CCTGGATGAGTGGGAGCGCGAGATCGACAACACTACCAACCTGTATCTACACCTGATCG 1678

QY 419 TGCAGAGCAATAAGTCGGGAGCAGAACTTGCAGGACGACCTGGATTGGTACAGCGGC 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1679 AGGAGAGCCAGAACACGAGGAGAAAGACGAGCAGAGCTGCTGGAGCTGGACAAGTGGG 1738

QY 479 TGGTGTGAGCCCTTCTGGAGGGCGAGCCGACATCTCCGGGGCGCCATCACTTCAAGCA 538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1739 CCAGCCTGTGGAACCTGTTTCGACATCAGCAAGTGGCTGTGGTATACATCAAGATCTTCATCA 1798

QY 539 CCGATTGCTGTCGACCGGCCCGGACAGTCTTCTCCAGGCCACCGCGGAGAGGCC 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1799 TGATCGTGGCGGCCCTGGTGGGCCCTGCGCATCGTGTTCACCGTGTGAGCATCGTGAACC 1858

QY 599 GCATCTGCTCCAAGACCTGTCTCTCCGACACCCACCTGGCCACGCACTCTCTGGAGA 658
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1859 GCGTGGCCAGGGCTACAGCCCGCTGAGCTTCCAGACCGCGCTTCCCGCCCGCCCGCGGCC 1918
```

Search completed: February 13, 2003, 00:16:20  
Job time : 246 secs

QY 659 CCGAGTGGTTCCACGCGCTCCGCGCAAGTGGAGGCCCCCACTTACACCGCCGCGGCCCA 718  
| | | | |  
Db 1919 CCGACCGCCCGAGGAGGATCGAGGAGGAGGCGCGGCGGACCGCGAGCCAGCAGCC 1978  
| | | | |  
QY 719 ATCAGGGGCCCCGGGCGCTGGGCCACAGCTGGGGGCGCAAGGACGGGCTCGGGGGGACA 778  
| | | | |  
Db 1979 CCCTGGTGACGGCCCTGCTGGCCCTGATCTGGAGACACCTGGCGAGCCTGTGCCTTTCA 2038  
| | | | |  
QY 779 AGAGCCACTTCAAGTG 794  
| | | | |  
Db 2039 GCTACCACCGCCTGCG 2054  
| | | | |

RESULT 15  
US-09-476-242-8  
; Sequence 8, Application US/09476242  
; Patent No. US20020146683A1  
; GENERAL INFORMATION:  
; APPLICANT: BARNETT, Susan  
; APPLICANT: HARTOG, Karin  
; APPLICANT: MARTIN, Eric  
; TITLE OF INVENTION: MODIFIED HIV ENV POLYPEPTIDES  
; FILE REFERENCE: 1605.002  
; CURRENT APPLICATION NUMBER: US/09/476,242  
; CURRENT FILING DATE: 1999-12-30  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 2316  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Val120-Thr202  
US-09-476-242-8

Query Match 1.2%; Score 45.6; DB 10; Length 2316;  
Best Local Similarity 44.0%; Pred. No. 0.024;  
Matches 192; Conservative 0; Mismatches 244; Indels 0; Gaps 0;  
QY 359 CPTCCTTGACCGGCGCTGGACACTGACACACGCCCACTTCCTCAACGGTATGC 418  
| | | | |  
Db 1619 CTTGGATGGAGTGGAGCGCGAGATCGACAACCTACACCACTGTATACACCTGTATCG 1678  
| | | | |  
QY 419 TGCAGAGCAATAAGTCGCGGGAGCAGAACTTGCAGGACGACCTGGATTTGGTACAGGCGC 478  
| | | | |  
Db 1679 AGGAGCCAGAACACAGCAGGAGAAAGACGAGCAGGAGCTGTGGAGCTGGACAAGTGG 1738  
| | | | |  
QY 479 TGGTGGAGCCCTTCGGAGGGGCGGCCAGCATCTCCGGGGCGCCATCACCTTCAGCA 538  
| | | | |  
Db 1739 CCAGCCTGTGGAACCTGTTTCGACATCAGCAAGTGGCTGTGTATACATCAAGATCTTCATCA 1798  
| | | | |  
QY 539 CCGATTTCGTGTCGCGACCGCGCCACAGGTCTTCCTCCAGGCCACCGCGGAGGAGCC 598  
| | | | |  
Db 1799 TGATCTGGGCGCCTGTGGGGCCTCGGATCGTGTTCACCGTGTGAGCATCGTGAACC 1858  
| | | | |  
QY 599 GCATCTGTCTCAAGACCTGTCTCTCCGCGACCCCACTGGCCAAACGCACTCTGGAGA 658  
| | | | |  
Db 1859 GCGTGGCCAGGCTACAGCCCTCTGAGCTTCAGACCCGCTTCCCGGCCCGCGCGCC 1918  
| | | | |  
QY 659 CCGAGTGGTTCCACGCGCTCCGGCGCAAGTGGAGGCCCACTTACACCGCCGCGGCCCA 718  
| | | | |  
Db 1919 CCGACCGCCCGAGGCGCATCGAGGAGGAGGCGGCGGAGCGACCGAGCCAGCAGCC 1978  
| | | | |  
QY 719 ATCAGGGGCCCCGGGCGCTGGGCCACAGCTGGCGGCGCAAGGACGGGCTCGGGGGGACA 778  
| | | | |  
Db 1979 CCCTGGTGACCGCCCTGCTGGCCCTGATCTGGAGACACCTGGCGAGCCTGTGCCTTTCA 2038  
| | | | |  
QY 779 AGAGCCACTTCAAGTG 794  
| | | | |  
Db 2039 GCTACCACCGCCTGCG 2054  
| | | | |



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2003, 04:23:28 ; Search time 62.2905 Seconds  
(without alignments)  
2599.103 Million cell updates/sec

Title: US-09-775-181-2  
Perfect score: 6382  
Sequence: 1 MGAWAYPLLCLLLLAQLGLG.....LSANKIAGPKKEIWDSEFKV 1215

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6382	100.0	1215	22 AAE06642	Human G-protein co
2	3119.5	48.9	599	22 AAE06643	Human G-protein co
3	571.5	9.0	669	22 ABB70858	Drosophila melanog
4	571.5	9.0	669	22 ABB38983	Drosophila G-prote
5	307	4.8	1677	22 ABB69976	Drosophila melanog
6	281	4.4	264	22 ABB67184	Drosophila melanog
7	254	4.0	1221	22 ABB62615	Drosophila melanog
8	254	4.0	1221	22 ABB38942	Drosophila G-prote
9	243.5	3.8	1220	22 AAB86160	D. melanogaster GA
10	233.5	3.7	176	22 ABB67185	Drosophila melanog

11	232	3.6	1305	22 AAB86161	D. melanogaster GA
12	227.5	3.6	1305	22 ABB59497	Drosophila melanog
13	227.5	3.6	1305	22 AAU38927	Drosophila G-prote
14	223.5	3.5	1199	13 AAR25080	GLU-G-R subtype la
15	220	3.4	1219	18 AAR25763	Amino acid sequenc
16	212	3.3	1197	20 AAY28562	Mouse metabotropic
17	210	3.3	1171	20 AAY28563	Rat metabotropic g
18	205.5	3.2	1111	22 ABB60327	Drosophila melanog
19	205	3.2	1183	22 ABB58769	Drosophila melanog
20	204	3.2	1180	22 AAE01156	Human metabotropic
21	198.5	3.1	1812	22 ABB58022	Drosophila melanog
22	198	3.1	1180	16 AAR64253	Human mGluR5a. Ho
23	198	3.1	1397	20 AAY49134	pmGluR2/Car*Galpha
24	198	3.1	1397	23 AAO15100	Human pmGluR2-Ca*
25	197.5	3.1	1212	20 AAY28564	Human metabotropic
26	197.5	3.1	1212	22 AAB47220	hmGluR5b. Homo sa
27	197.5	3.1	1212	22 AAE01157	Human metabotropic
28	197.5	3.1	1402	23 AAO15105	Human ph2SPMGLUR3-
29	194	3.0	1043	22 AAB47219	Chimeric receptor
30	193.5	3.0	1212	16 AAR64254	Human mGluR5b. Ho
31	193.5	3.0	1394	20 AAY49129	pmGluR2/Car*Galpha
32	193.5	3.0	1394	23 AAO15095	Human pmGluR2-Ca*
33	193	3.0	1194	14 AAR42199	HSmGluR1. Homo sa
34	193	3.0	1422	23 AAO15102	Human phmGluR4-Ca*
35	193	3.0	1481	22 ABB69279	Drosophila melanog
36	190	3.0	1056	18 AAW25765	Amino acid sequenc
37	189	3.0	2194	22 AAM40114	Human polypeptide
38	188.5	3.0	1325	22 ABB29725	Peptide #2376 enco
39	188.5	3.0	1325	22 ABB34897	Peptide #2403 enco
40	188.5	3.0	1325	22 ABB20314	Protein #2313 enco
41	188.5	3.0	1325	22 AAM55707	Human brain marrow
42	188.5	3.0	1325	22 AAM68085	Human bone marrow
43	188.5	3.0	1325	22 AAM03645	Peptide #2377 enco
44	188.5	3.0	1325	23 ABB37612	Human peptide enco
45	188	2.9	1072	23 ABB54963	Lactococcus lactis

#### ALIGNMENTS

RESULT 1  
AAE06642  
ID AAE06642 standard; Protein; 1215 AA.  
XX AC AAE06642;  
XX AC  
DF 16-OCT-2001 (first entry)  
XX DE Human G-protein coupled receptor (NGPCR) #1.  
XX DE

Human: G-protein coupled receptor; NGPCR; gene therapy; cancer; diabetes;  
inflammation; immune disorder; heart disease; obesity; coronary disease;  
metabolic disorder; physiological disorder; therapeutic; drug screening;  
signal transduction; behavioural disorder.

OS Homo sapiens.  
XX PN WO200157086-A2.  
XX PD 09-AUG-2001.  
XX PF 01-FEB-2001; 2001WO-US03648.  
XX PR 04-FEB-2000; 2000US-0180414.  
XX PA (LEXI-) LEXICON GENETICS INC.  
XX PI Donoho G, Hilbun E;  
XX DR WPI; 2001-488870/53.  
XX DR N-FSDB; AAD12292.  
XX PT Novel isolated polynucleotides encoding human G protein coupled

receptor (NGPCR), useful as probe and for treating disease involving GPCR, detecting mutant NGPCR or abnormally expressed NGPCRs for diagnosing disease -

Claim 2; Page 66-69; 73pp; English.

The present sequence is human G-protein coupled receptor (NGPCR) which is a 7 transmembrane protein of G-protein coupled receptor (GPCR) super family. NGPCR is expressed in hypothalamus, foetal brain, brain, cerebellum and testis. NGPCR spans the cellular membrane and is involved in signal transduction pathways. Human NGPCR is useful for treating obesity, inflammation, immune disorders, diabetes, heart and coronary disease, metabolic disorders and cancer. NGPCR's are also used as reagents in assays for screening compounds that are useful in the therapeutic treatment of physiological and behavioural disorders. NGPCR is useful for identifying compounds that modulate NGPCR gene expression and/or gene product activity. NGPCR DNA is also useful for the diagnostic and prognostic evaluation of disorders related to NGPCR function. Nucleotide constructs encoding functional NGPCR and mutant NGPCR variants are used in gene therapy. NGPCR probes are also useful for the identification, selection and validation of novel molecular targets for drug discovery.

Query Match 100.0%; Score 6382; DB 22; Length 1215;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1215; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGAMAYPLLLCLLQAOLGLGAGASRDQGRDPSRPTPKGKHAQOPGRASDSSAP 60  
 DB 1 MGAMAYPLLLCLLQAOLGLGAGASRDQGRDPSRPTPKGKHAQOPGRASDSSAP 60

QY 61 WSRSDTGITLAQKLAEEVPMVASYLYTGDSHQLKRANCSGRYELAGLPCKWPALASAH 120  
 DB 61 WSRSDTGITLAQKLAEEVPMVASYLYTGDSHQLKRANCSGRYELAGLPCKWPALASAH 120

QY 121 SLHRALDTLTHTATNLFNVLQSNKREONLQDDLDWYQALVMSLLEGEPPSISRAAITFT 180  
 DB 121 SLHRALDTLTHTATNLFNVLQSNKREONLQDDLDWYQALVMSLLEGEPPSISRAAITFT 180

QY 181 DLSAPAPQVFLQATRESRILLQDLSSAPHLANATLETWFHGLRKRWRPHLHRRGN 240  
 DB 181 DLSAPAPQVFLQATRESRILLQDLSSAPHLANATLETWFHGLRKRWRPHLHRRGN 240

QY 241 QGPRGLGHSWRKDGGLGDKSHFKWSPYLECENGSYKPGWLVTLSAIYGLQPNLVPEF 300  
 DB 241 QGPRGLGHSWRKDGGLGDKSHFKWSPYLECENGSYKPGWLVTLSAIYGLQPNLVPEF 300

QY 301 RGVMKVDINLQKVDIDQCSSDGFSGTHKCHLNNECMPIKGLGFVLGAYECICKAGFYH 360  
 DB 301 RGVMKVDINLQKVDIDQCSSDGFSGTHKCHLNNECMPIKGLGFVLGAYECICKAGFYH 360

QY 361 PGVLPVNNFRRRGPDQHSIGSTKDVSEBAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420  
 DB 361 PGVLPVNNFRRRGPDQHSIGSTKDVSEBAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420

QY 421 ISFOGLCMLDFVSMVYHFRKAKSIRASGLILLETILFGSLLLYFPVWILYFEPSTFR 480  
 DB 421 ISFOGLCMLDFVSMVYHFRKAKSIRASGLILLETILFGSLLLYFPVWILYFEPSTFR 480

QY 481 CILLRWALLGFATYVGTWTLKLRVLKVFLSRTAQRIPYMTGGVRMRLAVILLVWFVF 540  
 DB 481 CILLRWALLGFATYVGTWTLKLRVLKVFLSRTAQRIPYMTGGVRMRLAVILLVWFVF 540

QY 541 LGWTSSVCQNLEKQISLIGQCKTSDHILFNCLIDRWDMYTAFAEFLFLWGVYLCYAV 600  
 DB 541 LGWTSSVCQNLEKQISLIGQCKTSDHILFNCLIDRWDMYTAFAEFLFLWGVYLCYAV 600

QY 601 RTVPSAFHEPRYMAVAVINELIISAFITIRFVLASRLQSDWMLMLYFAHTHLTVTVIG 660  
 DB 601 RTVPSAFHEPRYMAVAVINELIISAFITIRFVLASRLQSDWMLMLYFAHTHLTVTVIG 660

QY 661 LLLIPKFSHSSNNPRDDIATAYEDELDMRGSGSYLNSINSAMSEHSILDPEDIRDELKK 720  
 DB 661 LLLIPKFSHSSNNPRDDIATAYEDELDMRGSGSYLNSINSAMSEHSILDPEDIRDELKK 720

QY 721 LYAQLEIYKRKKMTNNPHLQKRCSSKGLGRSIRMRITEIPETVSQCSKEDKEGADHG 780  
 DB 721 LYAQLEIYKRKKMTNNPHLQKRCSSKGLGRSIRMRITEIPETVSQCSKEDKEGADHG 780

QY 781 TAKGTALIRKKNPPESNGTNGSKKEETLKNRVFSLKKSHTYDHRVDQTEESSSLPESQE 840  
 DB 781 TAKGTALIRKKNPPESNGTNGSKKEETLKNRVFSLKKSHTYDHRVDQTEESSSLPESQE 840

QY 841 EETTENSTLESLSGKKLTQKLKEDSEAESESTESVPLVCKSASAHNLSSEKKTGHPTSMQ 900  
 DB 841 EETTENSTLESLSGKKLTQKLKEDSEAESESTESVPLVCKSASAHNLSSEKKTGHPTSMQ 900

QY 901 KSLSVIASAKETLGLAGKTQTAGVEERTKSQKPLPKDKETNRNHSNDNTETKDPAPON 960  
 DB 901 KSLSVIASAKETLGLAGKTQTAGVEERTKSQKPLPKDKETNRNHSNDNTETKDPAPON 960

QY 961 SNPAEPRKPKQSGIMKQORVNPPTTANSLAPGTTQMKDNFDIGEVCPWEYDLPVGPVP 1020  
 DB 961 SNPAEPRKPKQSGIMKQORVNPPTTANSLAPGTTQMKDNFDIGEVCPWEYDLPVGPVP 1020

QY 1021 SESKQKHVSIVASEMEKNPTFSLKEKSHHKPKAAEVCQSQSNOKRIDKAEVCLWESQGS 1080  
 DB 1021 SESKQKHVSIVASEMEKNPTFSLKEKSHHKPKAAEVCQSQSNOKRIDKAEVCLWESQGS 1080

QY 1081 ILEDEKLLISKTVPVLPRAKEENGQGPRAANVCAGSQSEBELPPKAVASKTENENLQIGHQ 1140  
 DB 1081 ILEDEKLLISKTVPVLPRAKEENGQGPRAANVCAGSQSEBELPPKAVASKTENENLQIGHQ 1140

QY 1141 EKKTSSESENVGSGSYNSNNFQQPLTSRAEYCPWEFETPAQPNAGRSVALPASSALSANK 1200  
 DB 1141 EKKTSSESENVGSGSYNSNNFQQPLTSRAEYCPWEFETPAQPNAGRSVALPASSALSANK 1200

QY 1201 IAGPRKEEINDSFVK 1215  
 DB 1201 IAGPRKEEINDSFVK 1215

RESULT 2  
 AAEO6643  
 ID AAEO6643 standard; Protein; 599 AA.  
 XX  
 AC AAEO6643;  
 XX  
 DT 16-OCT-2001 (first entry)  
 XX  
 DE Human G-protein coupled receptor (NGPCR) #2.  
 XX  
 KW Human; G-protein coupled receptor; NGPCR; gene therapy; cancer; diabetes; inflammation; immune disorder; heart disease; obesity; coronary disease; metabolic disorder; physiological disorder; therapeutic; drug screening; signal transduction; behavioural disorder.  
 KW  
 OS Homo sapiens.  
 XX  
 FN WO200157086-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 01-FEB-2001; 2001WO-US03648.  
 XX  
 PR 04-FEB-2000; 2000US-0180414.  
 XX  
 PA (LEXI-) LEXICON GENETICS INC.  
 XX  
 PI Donoho G, Hilbun E;  
 XX  
 DR WPI; 2001-488870/53.  
 DR N-PSDB; AAD12293.  
 XX



PT Novel isolated polynucleotides encoding human G protein coupled  
PT receptor (NGPCR), useful as probe and for treating disease involving  
PT GPCR, detecting mutant NGPCR or abnormally expressed NGPCRs for  
PT diagnosing disease -  
XX  
PS Claim 4; Page 70-71; 73pp; English.  
XX  
CC The present sequence is human G-protein coupled receptor (NGPCR) which  
CC is a 7 transmembrane protein of G-protein coupled receptor (GPCR) super  
CC family. NGPCR is expressed in hypothalamus, foetal brain, brain,  
CC cerebellum and testis. NGPCR spans the cellular membrane and is involved  
CC in signal transduction pathways. Human NGPCR is useful for treating  
CC obesity, inflammation, immune disorders, diabetes, heart and coronary  
CC disease, metabolic disorders and cancer. NGPCR's are also used as  
CC reagents in assays for screening compounds that are useful in the  
CC therapeutic treatment of physiological and behavioural disorders. NGPCR  
CC is useful for identifying compounds that modulate NGPCR gene expression  
CC and/or gene product activity. NGPCR DNA is also useful for the diagnostic  
CC and prognostic evaluation of disorders related to NGPCR function.  
CC Nucleotide constructs encoding functional NGPCR and mutant NGPCR variants  
CC are used in gene therapy. NGPCR probes are also useful for the  
CC identification, selection and validation of novel molecular targets for  
CC drug discovery.

XX Sequence 599 AA;  
SQ  
Query Match 48.9%; Score 3119.5; DB 22; Length 599;  
Best Local Similarity 98.5%; Pred. No. 1.3e-220;  
Matches 585; Conservative 2; Mismatches 2; Indels 5; Gaps 1;  
QY 1 MGAMAYPLLLCLLLAQLGAGVASRDPOGRPDSPRTPKGPQAQPGRASADSSAP 60  
Db 1 MGAMAYPLLLCLLLAQLGAGVASRDPOGRPDSPRTPKGPQAQPGRASADSSAP 60  
QY 61 WSRSTDTTILAQKLAIEVPMVASIYLTGDSHQLKRANCSGRYELAGLPGKWPALASAH 120  
Db 61 WSRSTDTTILAQKLAIEVPMVASIYLTGDSHQLKRANCSGRYELAGLPGKWPALASAH 120  
QY 121 SLHRALDITLTATNLFNVLQSNKREONLQDDLDWYQALVMSLLEGEPSISRAAITFTST 180  
Db 121 SLHRALDITLTATNLFNVLQSNKREONLQDDLDWYQALVMSLLEGEPSISRAAITFTST 180  
QY 181 DLSAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPHLHRRGPN 240  
Db 181 DLSAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPHLHRRGPN 240  
QY 241 QGPRGLGHSWRRKDGKSHFKWSPPYLECEGSKYKPGWLVTLSAIYGLQPNLVPEF 300  
Db 241 QGPRGLGHSWRRKDGKSHFKWSPPYLECEGSKYKPGWLVTLSAIYGLQPNLVPEF 300  
QY 301 RGVKMYDINLQKVIDIDQCSSDGFSGTHKCHLNNSCMPKIGLGFVIGAYECICRAGFYH 360  
Db 301 RGVKMYDINLQKVIDIDQCSSDGFSGTHKCHLNNSCMPKIGLGFVIGAYECICRAGFYH 360  
QY 361 PGVLPVNNFRRGPDQHIISGSKVDSEAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420  
Db 361 PGVLPVNNFRRGPDQHIISGSKVDSEAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420  
QY 421 ISFOGLCMLLDFVSMVLVYVHFRKAKSIRASGULLILETLFGSLLILYFPVILYFEPSTFR 480  
Db 421 ISFOGLCMLLDFVSMVLVYVHFRKAKSIRASGULLILETLFGSLLILYFPVILYFEPSTFR 480  
QY 481 CILLRWARLLIGFATVYGTVTLKHLRVKVLFSRTAQRIPTYMTGGVRMRLAVTLLVWF 540  
Db 481 CILLRWARLLIGFATVYGTVTLKHLRVKVLFSRTAQRIPTYMTGGVRMRLAVTLLVWF 540  
QY 541 LIGWTSVQCNLEKQISLIGQGTSDHLLIFNMCLIDRWDMYTAFAEFLFLWVG 594  
Db 541 LIGWTSVQCNLEKQISLIGQGTSDHLLIFNMCLIDRWDMYTAFAEFLFLWVG 594  
RESULT 3  
ABB70858

ID ABB70858 standard; Protein; 669 AA.  
XX AC ABB70858;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 39366.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PR 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EW;  
XX DR WPI: 2001-656860/75.  
XX DR N-PSDB; ABL14961.  
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX PS Disclosure; SEQ ID NO 39366; 21pp + Sequence Listing; English.  
XX CC The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).  
XX CC The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 669 AA;  
Query Match 9.0%; Score 571.5; DB 22; Length 669;  
Best Local Similarity 26.1%; Pred. No. 4e-33;  
Matches 195; Conservative 113; Mismatches 316; Indels 123; Gaps 27;  
QY 110 GKWPALASAHPSLHRALDITLTATNLFNVLQSNKREONLQDDLDWYQALVMSLLEGE 169  
Db 6 GKWPC-RMEFYDITDAEDAARQFIEFL-----SGKFPNANTPIAID-----EP 47  
QY 170 SISRAAITFSTLSAPAPQVFLQATREESRILLQDLSAPHLA-----NATLETWF 223  
Db 48 --TRAEVSRANGIAS-----YALNEDDNLATAAPSIHTVVVKFRDNVTIPPDQV 98  
QY 224 HG---LRRKWRPHLHRRGPNQGRGLGHSWRRKDGKSHFKWSPPYLECEGSKYKPG 280  
Db 99 HNKAYLGSWRE-----LGAWNSTDGTQ-----EWGAPRDCNLLTRRWL 139  
QY 281 WLVTLSAIYGLQPNLVPEFRGMKVDINLQKVIDIDQCSSDGFSGTHKCHLNNSC 337  
Db 140 WFERISFS-----EHR-IKVVAAAFIAADEVC-NDGLEEVFGRHGCDDRTTFC 187  
QY 338 MPKIKGL-GFVLGAYECICRAGFYHPCVLPVNNFRRGPDQHIISGSKVDSE--EAYVCLP 394  
Db 188 LLTENKPAATRDVYTCLCRESYLPN-STLQGR-----GDRVELSEGYDNYSCIP 237





Db 1 MYPNPNLTCTARIWLREIGFSLTYGALMLKTWRISVIFRVSRAKAV-ITDAALLKRLG 59

QY 532 VILLVVFELGWTSVQCNLEKQSL--IGOGKTSDLHIFNCLIDRWYMTAVAEFLF 589

Db 60 IICGAI-----GTCLLVFTLVSPDVVGVRTADDKAFCKCTDMWDYTTFTSMEVLF 110

QY 590 LLWGVYLCYAVRTVPSAFHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLVFA 649

Db 111 LANGVRLCITWRKAPSEFNESRFISMAIYNEFLCFLVSMFLFOSPANPDLLIIFPC 170

QY 650 HPHLFTVTVIGLL---IPKFSHSSNNPRDDIATEAYEDELDMGRSGS 694

Db 171 HQLTAVTLLALIFGSKIPTLQGVIVLR---SGKSHQENIGMGTKAS 215

RESULT 7

AB062615

ITD AB062615 standard; Protein; 1221 AA.

AB062615;

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 14637.

Drosophila; developmental biology; cell signalling; insecticide;

pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US09231.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-0614150.

(PEKE ) PE CORP NY.

Venter JC, Adams M, Li PWD, Myers EW;

WPI; 2001-656860/75.

N-PSDB; ABL06718.

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Disclosure; SEQ ID NO 14637; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1221 AA;

Query Match

Best Local Similarity 4.0%; Score 254; DB 22; Length 1221;

Matches 213; Conservative 195; Mismatches 496; Indels 288; Gaps 46;

QY 72 OKLAEEVPMVASYLYTGDGHQKLRANGSG-RYELAGLPGKWPALASAHPSLHRALDTLT 130

Db 216 KKLREK---DVRILGNFNEHFARFAFCEAYKLDWYGRAYQWLIMA-----T 259

QY 131 HATNFLNVMLOSKSREONLQDDLDWYQALVWSLLEGEPSISRAAITFTSDLSAPAPQV 190

Db 260 YSTDWNTQDSECVER-----IATALEG-----AILVDLLPLSTSGDIT 300

QY 191 FLQATREESRILLQDSSAPHLANATLETFWFHGLRRK--WRPHLHRRGPNQGRGLGH 248

Db 301 VAGITADEVLEYDLRGT-----EYSRFHGYTVDGIWAAALAIOYVAE----- 344

QY 249 SWRRKDGIGGDKSHF-----KWSPPYLEC-ENGSKYKQWLVTLSAIYGLQNLVPEFRG 302

Db 345 --KREDLL---THFDYRVKDWESVLEALRNTSEF---VTGPVRFYNN----- 386

QY 303 VMKVDINLQKVDDIDOCSSDGFSGTHKCHLNNSCMPKIGLGLFVLGAYECICKAGFYHPG 362

Db 387 -RKANILINQFOLGOMEKIGEYH-SOKSHLDLSLQKPKVWVG----- 426

QY 363 VLPVNNFRRRGPDQHISGTDKDVSEAYVCLPCREGCPFCADDSPCFVQEDK-----YLR 417

Db 427 -----KTPPK-----DRTLIIYIEHSQVNFYIV 450

QY 418 LAISFQGLCMLLDVFSMLVYHFRKAKSIRASGLILLETILFGLSLLLYFPVILYFEP 477

Db 451 SASASVIGVIATVFLAFNIKY--RNQRYIKMSSPHLANLIIVGCMYTLSTIFLGLDT 508

QY 478 -----TFRCILLRWALIGFATVYGTVTLKLRHVKVLSRTAQRIPTYMTGGVMMRL 530

Db 509 LSSVAAPFYICTARAWILMAGFSLSGFAMFSTWRVHSIFTDLKLNK-KVIKDYQLFWV 567

QY 531 AVILLVWFVFLIGW-TSSVCQNLKQISLIGOCKTSDHLIFNMCLIDRWYMTAVAEFLF 589

Db 568 GVLLAIDIAITWQIADPFYRETQLEPLHHENDDLVIPENEVYCOSEHMTIFVSIY 627

QY 590 -----LLWGYLCYAVRTVP-SAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDW 642

Db 628 AYKGLLVFGAFLAWEYTRHVSIPALNDSKHIGFSYVNFITCLAGAAISLVSLSRDKLVF 687

QY 643 MLMLYFAHTLTVTIGLLIPKFSHSSNPRDDIATEAYEDELDMGRSGSYLANSINS 702

Db 688 VLLSFF--IIFCTATLCLVFPKLVELKRNPGVVDVRRVRLATLRPMKNGR----- 737

QY 703 AMSEHSLDPEDIRDELKLYAQLEIYKRKKMITN-NPHLQKKRCKSKGIGRSIMRRIETE 761

Db 738 -----RDS-----SVCELEQRLRDVKNCRFRKALMEKENELQALIRKLG-- 778

QY 762 PETVSRQ-----CSKDEKADHGTAKGTALIRKNPSSSGNTCKSKEETLKNRVFSLK 815

Db 779 PE--ARKWIDGVTCTGGSNVSGSELEPILNDDIVRLSAPPVRRMPSTTEVMTSVDSVT 836

QY 816 KSHSTYDH---VRDQTESSSLPTESQEEETEN-----STLESLSGKKLTOKLKEDSE 866

Db 837 STHVEMDNFSVQSVTVMAPSLPPKKKQSVIHHSHAPATMMQPIQQQLQHQHQ 896

QY 867 AESTESVPLVCKSAHNLSEKKTGHPRTSMQLKSLSVIASAKEKTLGLAGKTOTAG-- 924

Db 897 MOOQH-----LQOOHQOQOQOQOQHHRHLEKRNVSVAQTD--NIGSIITACKR 948

QY 925 -----VEERTK-----SOKPLPKDKETNRNHSNDNTETKDPAPQNSPASEP 967

Db 949 SGGDCSSRRERRQSTASRHYDSGQTPTARPKYSSSHRNSNTISTQSSELSNMCHSKP 1008

QY 968 RKPQSGTLMKQORVNTTANSDLNPGTTQMKDNDFIGVCWPVEVYD-----TPGPVP 1020

Db 1009 STP---AVIK-----TPTASDHRTSMGSAKSNFVVSOSDLWDHTLTHAKQORQPRNYA 1061

QY 1021 SESKVQKHVSIVASEMEKNPT-----FSLKEKSHHKPKAAEVCCQSNQKRIDKAEVCLW 1074

Db 1062 SPQCAEHHGGHGTIDYDNTTSPIQRSVSEKRNKHKRPK-----QKGTVC-- 1107

QY 1075 ESQGSITLEDKLLISKTPVLPRAKEENGOPRAANVCAGQSEELPPKAVASKTENENL 1134

Db 1108 QSETDSERERPPPNQPCQVQPRKYS-----RSSNQHAHHHSSPNVADPK---QRS 1157





```
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL11288.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PT Disclosure: SEQ ID NO 28347; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 176 AA;
Query Match 3.7%; Score 238.5; DB 22; Length 176;
Best Local Similarity 31.5%; Pred. No. 1.7e-09;
Matches 52; Conservative 25; Mismatches 55; Indels 33; Gaps 4;
QY 301 RCVMKVDINLQKVDIDQSSD-----GWFSHTKCHLNNSCMPKIGLGFVLGAYE 351
DB 9 RGTSDIDILRRVDIDQCPQRHTPGTKRPLNIFAGTDKCKQRTMCEAIMGLGFRGSYK 68
QY 352 CIKAGFYHPGVLPVNNFRRRGPQDHISGTDKV-----SEAYVCLPCRE 397
DB 69 CLCRGFGYFPDIVSQHKP-----FNGSLLSEYKMLGKNSTYNSNSEYECPLPCAE 120
DB 398 GCPFCADSPCFVQEDKYLRALAIISFQGLCMLLDFVSMVLVYHFR 442
DB 121 GDCSDSDSPCIAALNWPMTSILAL--ACIVIGLLPPAAWFTFR 163
RESULT 11
AAB86161
ID AAB86161 standard; Protein; 1305 AA.
XX AC AAB86161;
XX AC AAB86161;
XX DT 03-AUG-2001 (first entry)
XX DE D. melanogaster GABA-B receptor protein SEQ ID 6.
XX KW GABA-B receptor; fruitfly; gamma-aminobutyric acid B receptor;
XX KW insecticide; transgenic invertebrate; plant protection agent;
XX KW human medicine; veterinary medicine; insect.
XX OS Drosophila melanogaster.
XX XX
XX PN DE19955408-A1.
XX PD 23-MAY-2001.
XX PF 18-NOV-1999; 99DE-1055408.
XX PR 18-NOV-1999; 99DE-1055408.
XX PA (FARB ) BAYER AG.
XX PI Raming K, Mezler M, Mueller T;
XX DR WPI; 2001-318282/34.
XX DR N-PSDB; AAH20521.
XX PT New invertebrate gamma-aminobutyric acid receptor proteins, useful in
XX PT screening for potential insecticides, for plant protection or medicine,
XX PT also related nucleic acid
XX PS Claim 2; Page 51-59; 62pp; German.
XX CC This invention describes a novel polypeptide (I), functioning as a
XX CC gamma-aminobutyric acid -B (GABA-B) receptor which has insecticidal
XX CC activity. (I), also the nucleic acid (II) that encodes it and related
XX CC vectors, host cells, antibodies and transgenic invertebrates, are used
XX CC for identifying: (i) new plant protection agents, i.e. modulators of (I)
XX CC with insecticidal activity, which may also be useful in human or
XX CC veterinary medicine; and (ii) genes that encode polypeptides involved in
XX CC assembly of functionally related GABA-B receptors in insects. This
XX CC sequence represents a fruitfly (Drosophila melanogaster) GABA-B receptor
XX CC which is described in the method of the invention.
XX SQ Sequence 1305 AA;
Query Match 3.6%; Score 232; DB 22; Length 1305;
Best Local Similarity 18.6%; Pred. No. 1e-07;
Matches 228; Conservative 183; Mismatches 396; Indels 416; Gaps 59;
QY 63 RSTGTTLAQKLAEEVPMDVASYLYTGDShOLKCRANCSGR-----YELAGLPGKWP--A 114
DB 344 RETDTRIIGSFQELAPQILCEAY-----RLRMFGADYAWILHESMGAPW-WPQQR 394
QY 115 LASAHPSLHRLDTL---TH-----ATNFLNMVLSQNSKREONLQDDLDWQALVW 162
DB 395 TACSNEHLOLAVENTLIVVSTHNSIVGNVSYGLNMFNSOLRKQSAQ----- 443
QY 163 SILGEPSSIRAAITFTSDLSAPAPQVFLQATREESR-----ILIQDLSS 208
DB 444 --FHGDCFG-----SGYCPRIASIAOTSDSRRRRRGVVGTSGGHLPEALISQ 490
QY 209 SAPHLANATLETWFHGLRRKWRPHLHRRGNQGRGLGHSWRKRDGLGDKSHFKWSPP 268
DB 491 YAPQ---TYDAVWAIALALRAAEHWRNEEQS-----KLDGFDYTRSDMAWE-- 535
QY 269 YLE-----CENGSKPGWLVTLSAATYGLQPNLVPFRCGMKVDINLQKVD 314
DB 536 FLQOMGKLHFLGVSGPVSFSGPDRVG-----TTAFYIQ-----LEPVA 577
QY 315 IDQCSSDGFSGTHKCHLNNSCMPKIGLGFVLGAYECICKAGFYHPGVLPVNN--FRRR 372
DB 578 L-----YYPATDALDFRCPCRPVK-----WHSQVPIAKRVFKLR 613
QY 373 GPDQHISGTDKDVSEEAIVCLPREGCPFCADDSPCFVQEDKYLRALAIISFQGLCMLLDF 432
DB 614 -----VATIAPL-----AFVTIATLSSVGIATAITF 639
QY 433 VSMVLVYHFRKAKSIRASGLILLETILFGSLLLPVFPVILYFEPST-----FRCILLR 485
DB 640 LAFNL--HFRKLKAIKLSPPKLSNITAVGCCIFVYATVILLGDHSTLPSAEDSFATVCTA 697
QY 486 WARLL--CFATVYGVTVTLKRLVKVFLSRTAQ-----RIPYMTGVRMRLAVI 533
DB 698 RYLLSAGFSLAFGSMFAKTYRVHRIF--TRTGSVFKDKMLQDIQLILLVG---LLLVDA 753
QY 534 LLVVFVFLGWTSSVCONLEKQISLIGOKTSDHLIF-----NMCL---IDRWYNTAVAE 586
DB 534 LLVVFVFLGWTSSVCONLEKQISLIGOKTSDHLIF-----NMCL---IDRWYNTAVAE 586
```

Db 754 LLVTLWVYDTPMERHLNLTLEIS-----ATDRSVVYQVQVEVCRSQHTQTWLSVLYAYK 808  
 Qy 587 FLFLWGLVYLCVAVTVP-SAFHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLM 645  
 Db 809 GULLVGVYMAWETHRVKIPALNDSQYIGVSVYVITSAI-----VVLANLISERVTL 863  
 Qy 646 LYFAHTL---TFTVTIGLLIPKFSHSSNPRDDIATEAYEDELDMGRSGYLSNINS 702  
 Db 864 AFITITALTSTTATCLLPK-----LHD 890  
 Qy 703 ANSEHS-LDP-----EDIRDELKLYAQLEIYKRMKMTNPHLOKK 743  
 Db 891 IWARNDIIDPVTHSMGLKMECNTRFVVDRELRQYRVEQVRVYKKE-----IQAL 942  
 Qy 744 RSKKGLGRSIRMRITETPEVSRQCKEDKEGADH-----GTAKGTALIRKNPPE 794  
 Db 943 DAETIKRLERLESGLTSTTTTSSSTSL--ITGGGHLKPELTVTSGISOTPAASKNRTPS 1000  
 Qy 795 SSGNTGKSKEETLKNRVSFL-----KKSHSTYDHRVDQTEESSLPTESEOEETTENST 848  
 Db 1001 TSG-----ILPNLLLSVLPPVTPRASWPSAEYMQIPMRSSVTFASQPOLEE----- 1046  
 Qy 849 LESLSGKKLTOKLEDSEASTESVPLVCKSAS-----AHNLSSEKKTG-----H 893  
 Db 1047 -----ACLPAAQDLINLRLAHQQAQATEAKTGLINLRGIF 1079  
 Qy 894 PRTSMLOKLSLVASAKETLGLAGKTOTAGVEERTKSQKPLPKDKETNRHNSDNTET 953  
 Db 1080 SRTTSSNGK-STASLADQKGKAAFKSH-MGLFTPL-----IPSSQTASCAIYNPNQD 1132  
 Qy 954 KDPAPQNSPABEPRKPKQSGIMKQQRVNTTANSDLNPGTQMKD-NF-DIGEVCPWEV 1011  
 Db 1133 SIPSEASSHPNGHLKPIHRGSL-----TKSGTHLDHLT---KDPNPLPIPTISGGEQ 1182  
 Qy 1012 YDLTPGPVPSESKVQKHVSIIVASEM-----EKNPTF-----SLKEKSHHKPK-----AA 1055  
 Db 1183 GDQTLG-----GKYVKLLETQYVNFQPSNRPRSVVQQPPSLRERVGRGSPRPHRLP 1234  
 Qy 1056 EVCQSNQKIDKAEVCLWESOGSILEDEKILLISKTPLVPERAKEE--NGGQPRRA----- 1109  
 Db 1235 PTCSLS-----ALAESDRP--GDSTSLGSKSIPRISLQQTSGTGWKSMETV 1282  
 Qy 1110 --ANVCAGQSEELPKAVASKTE 1130  
 Db 1283 GKSRLSLGDSQEEQAPANGTE 1305  
 JLT 12  
 ID ABB59497 standard; Protein: 1305 AA.  
 AC ABB59497;  
 XX  
 XX  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 5283.  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 OS Drosophila melanogaster.  
 XX  
 XX WO200171042-A2.  
 XX  
 XX  
 PD 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX

Venter JC, Adams M, Li PWD, Myers EW;  
 WPI: 2001-656860/75.  
 N-PSDB: ABL03600.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 5283; 2lpp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1305 AA;  
 Query Match 3.6%; Score 227.5; DB:22; Length 1305;  
 Best Local Similarity 18.8%; Pred. No. 2.2e-07;  
 Matches 209; Conservative 171; Mismatches 381; Indels 351; Gaps 52;  
 Qy 181 DLSAP-APQVFLQATREESRILLQDLSSSAPH-----LANATLETWPHGLRRKWRP 232  
 Db 383 ESMGAPWPDQRTACSNHELQLAVENTLIVSTHNSIVGNVSYSGLNHNHNSQLRKQSA 442  
 Qy 233 HLHRRGNQGRPLGHSW-----RRKDLGDKSHFKWSPYLCSENGSY 277  
 Db 443 QFH-----GQDGFSGYGSRIATAATQDSRRRRRGVGTGGHLPFEAISQYAPQTY 496  
 Qy 278 KPGWLVTLSAIYQLPNLPEFRGVKMDINLQKVIDOCSDGWFSGTHKCHLNSEC 337  
 Db 497 DAVMAIAL-----RAAEHW-----RRNEQ 519  
 Qy 338 MPKIGLGEVLG--AYEICKAGFYH-PGVLPVNNRRRGPDQHISGST-----KDYSEE 388  
 Db 520 SKLDGFDYTRSDMAWEFLQMGKHLFELGSGPVSF--SGPDR--VGTAFYQIQRGLLEP 575  
 Qy 389 AYVCLPCREG---CPEC-----ADDSPCFVQEDKYLRLAISEF 423  
 Db 576 VALYYPATDALDFRCPRCPVKHSGQVPIAKRVFKLRVATIAPL-----AFYTIATLS 630  
 Qy 424 QGLCMLLDYVSMVYHFRKAKSTRASGLLLETLFGLSLLXPPVVLVYFEPST----- 478  
 Db 631 VGIALATAFLAFNL--HFRKLKAIKLSPKLSNTAVGCFYVATVILLGLDHTLPSAE 688  
 Qy 479 --FRCILLRWARLL--GFATVYGTVTTLKHLRVLFVLSRTAQ-----RIPYMTGG 524  
 Db 689 DSPATVCTARVYLLSAGSLAFSGSMFAKTYVRHRIF--TRTGSVPFKMDLIQILLVGG 747  
 Qy 525 RVMRLAVILLVFWFLIGHTSSVCQNLKQISLIGOGKTSDDLIF---NMCL---IDR 577  
 Db 748 ---LLLVDAALLVTLWVYDTPMERHLNLTLEIS-----ATDRSVVYQVQVEVCRSQHTQ 799  
 Qy 578 WDMYTAVAEEFLWGLVYLCVAVTVP-SAFHEPRYMAVAVHNELIISAIFHTIRFVLAS 636  
 Db 800 WLSVLYAYKGLLVGVYMAWETHRVKIPALNDSQYIGVSVYVITSAI-----VVLAN 854  
 Qy 637 RLQSDWMLMLYFAHTL---TFTVTIGLLIPKFSHSSNPRDDIATEAYEDELDMGRSG 693  
 Db 855 NLISERVTLAFITITALTSTTATCLLPK----- 887  
 Qy 694 SYLNSINSINSAWSEHS-LDP-----EDIRDELKLYAQLEIYKRMKMTNPHLOKK 734  
 Db 888 -----LHDIWARNDIIDPVTHSMGLKMECNTRFVVDRELRQYRVEQVRVYKKE--- 938



QY 735 TNNPHLQKRCCKGKGRIMRRITETPTVSQCKEDKEGADH-----GTAKGT 785  
 Db 939 -----IQALDAEIRKLERLESGLTTTSSSTSL--LTGGHLKPELTVTSGISQPP 991  
 QY 786 ALIRKNPPSSGNTGSKKEETLKNRVFSL-----KKSHTYDHYRDQTEESSLPTESQ 839  
 Db 992 AASKNRTPSISG-----ILPNLLSVLPVPIPRASWPSAEYMOIPMRRSVTFASQPQ 1043  
 QY 840 EETTENTLESLSGKKLTQKLEDSEAESESTVPLVCKSAS-----AHLNLSPEKKTG 892  
 Db 1044 LEE-----ACLPADQLINLRLAHQOATEAKTG 1070  
 QY 893 -----HPRTSMLQKLSVTSASAKEKTLGAGKTQTAGVEERTKSQKPLPKDKETNRN 944  
 Db 1071 LINLRGIFSRITSSNKG-STASLADQKGLKAFFKSH-MGLFTRL-----IPSSQTASCN 1123  
 QY 945 HNSDNTETKDPAPONSFAEPRPKQSKIMKQQRVNPPTTANSDLNPGTTQMKD-NF-D 1002  
 Db 1124 AYNPNQDSIFSEASHPHNGNHLKPIHRGSL-----TKSGTHLDHLT---KDPNPLP 1173  
 QY 1003 IGEVCPWEYDILTPGPVPSKQKHSIVASEM-----EKNPTF-----SLKEKSHHK 1051  
 Db 1174 IPTISGEGDQTLG-----GKVKLLETQVNFQLPNSRRPSVVQPPSLRERVGS 1225  
 QY 1052 PK-----AAEVCOQSNQKIDKAECVLCWESQGSILEDEKLLISKTPVLPRAKEE--NG 1104  
 Db 1226 PRFPHRILPTCSLS-----ALAESEDRP--GDSTSLGSCSKSPRISLQOATSG 1273  
 QY 1105 GOPRA-----ANVCAGQSEELPPRAVASKTE 1130  
 Db 1274 GTWKMETAGKRLSLGDSQEBEQAPANGTE 1305  
 RESULT 13  
 ID AAU38927  
 XX AAU38927 standard; Protein; 1305 AA.  
 AC AAU38927;  
 DT 16-JAN-2002 (first entry)  
 XX Drosophila G-protein coupled receptor, GCPR #5.  
 DE Drosophila; G-protein coupled receptor; GCPR: insecticide; diagnostic;  
 KW mutation detection.  
 KW Drosophila melanogaster.  
 OS  
 XX WO200170980-A2.  
 PN 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09341.  
 PF 23-MAR-2000; 2000US-191638P.  
 PR 18-JUL-2000; 2000US-0618893.  
 XX (PEKE ) PE CORP NY.  
 PA Cravchik A;  
 XX WPI; 2001-616405/71.  
 DR N-PSDB; AAS57080, AAS57081.  
 XX Sixty six Drosophila Melanogaster G-protein coupled receptors (GPCR),  
 PT useful in the treatment and diagnosis of GPCR-related conditions and  
 PT for identifying GPCR modulators for use as insecticides -  
 XX Claim 1; Page 69-72; 392pp; English.  
 PS The invention relates to sixty six novel isolated Drosophila  
 CC melanogaster G-protein coupled receptors (GPCR). The GPCR proteins and  
 CC nucleic acids are useful in the treatment and diagnosis of GPCR-related

CC conditions. The GPCR proteins and nucleic acids are also useful for  
 CC identifying modulators of GPCR proteins for use as insecticides. The  
 CC nucleic acid can also be used to detect mutations in GPCR genes and gene  
 CC expression products such as mRNA. AAU38923-AAU38988 represent D.  
 CC melanogaster G-coupled protein receptor amino acid sequences of  
 CC the invention.  
 XX

SQ Sequence 1305 AA;

Query Match 3.6%; Score 227.5; DB 22; Length 1305;  
 Best Local Similarity 18.8%; Pred. No. 2.2e-07;  
 Matches 209; Conservative 171; Mismatches 381; Indels 351; Gaps 52;  
 QY 181 DLSAP-APQVFLQATREESRILLQDLSAPH-----LANATLETFEWHGLRRKWRP 232  
 Db 383 ESGAGPWPDPDORTACSNHELQLAVENLIVVSTHNSIVGNVSYSGLNHHMFNSQLRKQSA 442  
 QY 233 HUHRGPNQPGRLGHSW-----RRKDLGGDKSHFKWSPPPYLECENSY 277  
 Db 443 QFH-----GQDGFSGYGSRIISAATSDSRRRRRRGVGTSGGHLFPEAISQYAPQTY 496  
 QY 278 KPGWLVTLSAIYGLQPNLPFERGVKMDINLOKVDIDQCSDDGWFSGTHKCHLNNSEC 337  
 Db 497 DAWAIALAL-----RAAEHWH-----RRNEQ 519  
 QY 338 MPKIGLGVFLG--AYECICKAGFYH-PGVLPVNNFRRRGPQDHISGST-----KDVSEE 388  
 Db 520 SKLDGFDYTRSDMAWEFLOQMKLHFLGVSGPVSF--SGPDR--VGTTFAYIOIQRGLLEP 575  
 QY 389 AYVCLPREG-----CPFC-----ADDSPCFVQEDKYLRLAITSF 423  
 Db 576 VALYYPATDALDFCRPCRPVKWHSGQVP IAKRVFKLRVATIAPL-----AFYTIATLSL 630  
 QY 424 QGLCMLLDFVSMVYHFRKAKSIRASGLILLETILFGSLLLYFPVVIYFPEST----- 478  
 Db 631 VGIALATAFLAFLN--HFRKLAIKLSPKSNITAVGCFIVYATVILLGLDHLPSAE 688  
 QY 479 --FRGILLRWARLL--GFATVYGTVTLKLHRLVKFLSRTAQ-----RIPYMTGG 524  
 Db 689 DSFAVCTARVYLLSAGFSLAFGSMFAYTVRHRIF-FTGVSFVKDKMLQDILQLLVGG 747  
 QY 525 RVMRMAVILLVFWFLGWTSSVCQNKLEKISLGQKTSDDLIF-----NMLC---IDR 577  
 Db 748 ---LLVDALLVTLMVTDPMERHLNLTLEIS-----ATDRSVYVQVQVEVCRSOHTQ 799  
 QY 578 WDMYTAFAEFLFLNGVYLCYAVRTVP-SAFHEPRYMAVAVNELLISAIFHTIRFVLAS 636  
 Db 800 WLSVLYAYKGLLVGVYMAWETHRVKIPALNDSQYIGVSVSVVITSAI-----VWVLA 854  
 QY 637 RLQSDWMLMLYFAHTHL--TVTVTIGLLIPKFSHSSNNPRDDIATEAYEDELDMGRSG 693  
 Db 855 NLISERVIAFITITAILITSTATLCLLFIKP----- 887  
 QY 694 SYLNSINSANSEHS-LDP-----EDIRDELKKLYAQLEIYRKKMI 734  
 Db 888 -----LHDIWANDIIDPVHSMGLKMECNTRFRVDDRRRELQYRVEVQNRVYKE--- 938  
 QY 735 TNNPHLQKRCCKGKGRSIRRRITETPTVSQCKEDKEGADH-----GTAKGT 785  
 Db 939 -----IQALDAEIRKLERLESGLTTTSSSTSL--LTGGHLKPELTVTSGISQTP 991  
 QY 786 ALIRKNPPSSGNTGSKKEETLKNRVFSL-----KKSHTYDHYRDQTEESSLPTESQ 839  
 Db 992 AASKNRTPSISG-----ILPNLLSVLPVPIPRASWPSAEYMOIPMRRSVTFASQPQ 1043  
 QY 840 EETTENTLESLSGKKLTQKLEDSEAESESTVPLVCKSAS-----AHLNLSPEKKTG 892  
 Db 1044 LEE-----ACLPADQLINLRLAHQOATEAKTG 1070  
 QY 893 -----HPRTSMLQKLSVTSASAKEKTLGAGKTQTAGVEERTKSQKPLPKDKETNRN 944  
 Db 1071 LINLRGIFSRITSSNKG-STASLADQKGLKAFFKSH-MGLFTRL-----IPSSQTASCN 1123

QY 945 HNSDNTETKDPAPQNSPAEPRKPKQSGIMKQQRVNPNTTANSDLNPGCTQMKD-NF-D 1002  
Db 1124 AIYNNPNQDSIPSEASSHNGHLKPIHRGSL-----TKSGTHLDHLT---KOPNPLP 1173  
QY 1003 IGEVCPWEVYDTPGPVPSEKVKQHVSTVASEM-----EKNPTF-----SLKEKSHHK 1051  
Db 1174 IPTISGEGQDOTLG-----GRYVKLLKTKVNFQLPSNRRPSVVOQPPSLRERVRGS 1225  
QY 1052 PK-----AAEVCCQSQKQKIDRAEVCLEWESOGSILEDEKLLISKTPVLPERAKEE--NG 1104  
Db 1226 PRPPHILPPTCSLS-----ALASEDRP--GDSTSLGSCSKSIPRLSQQATSG 1273  
QY 1105 GOPRA-----ANVCAGQSEELPPKAVASKTE 1130  
Db 1274 GTWKSMTAGKSRLSLGDSQEERQAPANGTE 1305

RESULT 14  
25080  
AAR25080 standard; Protein; 1199 AA.  
XX AAR25080;  
AC AAR25080;  
KW Xenopus oocyte; cerebellum sublibrary; voltage clamp assay;  
KW neurotransmitter; glutamate; neuronal excitation; clone 45-A.  
XX Synthetic.  
OS  
FH Key Location/Qualifiers  
FT Modified-site 98 /note= "N-glycosylation"  
FT Modified-site 223 /note= "N-glycosylation"  
FT Modified-site 397 /note= "N-glycosylation"  
FT Modified-site 515 /note= "N-glycosylation"  
FT Domain 594..614 /label= Transmembrane\_domain  
FT Domain 630..650 /label= Transmembrane\_domain  
FT Domain 657..677 /label= Transmembrane\_domain  
FT Domain 707..727 /label= Transmembrane\_domain  
FT Modified-site 747 /note= "N-glycosylation"  
FT Domain 753..772 /label= Transmembrane\_domain  
FT Domain 787..807 /label= Transmembrane\_domain  
FT Domain 812..832 /label= Transmembrane\_domain  
FT Modified-site 920 /note= "N-glycosylation"  
FT Modified-site 925 /note= "N-glycosylation"  
FT Modified-site 1180 /note= "N-glycosylation"

XX WO9210583-A.  
PN  
XX 25-JUN-1992.  
PD  
XX 12-DEC-1991;  
PF 91WO-US09422.  
XX 12-DEC-1990;  
XX 30-JAN-1991;  
XX 18-MAR-1991;  
PR 90US-0626806.  
PR 91US-0648481.  
PR 91US-0672007.

XX (UNIW ) UNIV WASHINGTON.  
PA (ZYMO ) ZYMOGENETICS INC.  
XX  
XX Almers W, Hagen FS, Houamed KM, Mulvihill ER;  
PI WPI; 1992-234638/28.  
DR N-PSDB: AAQ25812.  
XX  
XX Mammalian G protein-coupled glutamate receptors - activate  
PT phospholipase C or stimulate inositol phospholipid metabolism,  
PT for use in diagnosis and identification of receptor agonists and  
PT antagonists  
XX  
XX Disclosure; Fig 5; 144pp; English.  
XX  
XX The sequence given is mammalian G protien-coupled glutamate  
CC receptor (GLU-G-R) subtype 1a. The clone encoding this protein was  
CC isolated from Xenopus oocyte cultures which had been injected with RNA  
CC extracted from a cerebellum sublibrary. Expression of GLU-G-R activity  
CC was assessed by voltage clamp assay. Plasmid DNA encoding GLU-G-R  
CC activity was isolated by replica-plating experiments. The protein  
CC encoded by clone 45-A was found to act as a receptor for the neuro-  
CC transmitter glutamate, and may play a critical role in the control of  
CC neurons. GLU-G-R and antibodies may be used to identify agonists and  
CC antagonists of GLU-G-R-mediated neuronal excitation and in diagnosis.  
XX  
SQ Sequence 1199 AA;

Query Match 3.5%; Score 223.5; DB 13; Length 1199;  
Best Local Similarity 18.6%; Pred. No. 3.8e-07;  
Matches 164; Conservative 121; Mismatches 325; Indels 271; Gaps 28;  
QY 357 GFYHPGVLPVNNFRRRRQDQHSSTKDVSEB-----AYCLPCREG- 398  
Db 498 GTWHEGVNLIDDDYKIQ---MNKSGMVRSCVSEPCLGKQIKVIRKGEVSCCWTACKENE 554  
QY 399 -----CPFC-----ADDSPCFQVEDKYLR-----LAIISFGQLCMLLDFFVSMLVV 438  
Db 555 FVDEFTCRACDLGWPNAELTGCEPIPVRYLEWSDIESIIAIAFSCGLGTLVFTVLTIF 614  
QY 439 YHFRKAKSIRASGLILLETILFGLSLLYFPVVILYFEPSTFCILLRWARLGFATVYGT 498  
Db 615 VLYRDPVVKSSRELVCVILAGIFLGVVCPPTLIAKPTTSCYLRQLLVGLSSAMCYS 674  
QY 499 VTLKLRVLKV-----FLSRTAQRIPYMTGGRVMRMLAVILLVVFVFLGW 544  
Db 675 LVTKTNRIARILAGSKKICIKTRKPRFMSAWAQVITIASILISVQLTFLVVTLLII----- 726  
QY 545 TSSVCQNLKQISLIGQKTSDDLIFNMCLIDRMVMTFAVA-EFLFLWGVVLCYAVRTV 603  
Db 727 -----MEPPMPLISYPSIKE--VYLICNTSNLCVAVPVGYNGLLIMSCYYAFTKRN 777  
QY 604 PSAPHEPRYMAVAVHNELIIISAIHFHTIRFVLASRLQSDWMLMLYFAHTHLVTVYTGILL 663  
Db 778 PANFNEAKYIAFTMYTTCIIWLAFPVIF-----GSNKKIITTCFAVSLSVTVALGCMF 831  
QY 664 IPK----FSHSSNPNRDDIATEAYEDELDMG-----RSGSYLNSINSINSAWSEHSLDP 711  
Db 832 TPKMYIIIAKPERNVRSFAFTT---SDVVRMHVGDGKLCPSRNTFLN----- 874  
QY 712 EDIRDELKLYAQLEIYKRRKMTNNPHLQKRCG-----KKGKGRSIRMRITPEIPET 764  
Db 875 -----IFRRKPGAGNANSNGKSVSWSPGGRQAPKGQHVWQORLSVHVKT 919  
QY 765 VSRQSKEDKEGADHTAGTALIRKNPPSSSGNTGKSKEETLKNRVFSLKSHSTYDHY 824  
Db 920 NETACNQ-----TAVIK-----PLKSY----- 937  
QY 825 RDQTESSSLPTSEEBETTENSTLESLSGKLLQKQKEDSEAESESTESVPLVCKSASAHN 884  
Db 938 -----QCGSNLTFSDASTKTLYNVEEDNTFTSAHFSPPSPSPSMVYHR 980

QY 885 LSSEKKTGHPRTSMLOKLSUSVTSASAKETLGLAGKTQTAGVBEERTKSQKPLPKDKETNRN 944  
Db 981 RGPVATTPPLPPLH-----TAEETPLFADSVIPKGL-----PPPLPQQQ----- 1021  
QY 945 HNSDNTETKDPAPQNSNPAERPRKQSGIMKQORVNPNTTANSDL-----NPG 993  
Db 1022 -----POQPPQPPQPPK--SLMDQLQGVVTFNGSGIPDFHAYLAGPGTGP 1066  
QY 994 TTQMKDNFDIGVCWPEVYDLTPGPVPS--SKVQKHVSIVASEMEKNPTFSLKCKSHHK 1051  
Db 1067 NT-----LRSLYPPPPPPHQLQMLPLHLSTFOEESISPPGEDIDDD-- 1108  
QY 1052 PRAAEVQSQSNOKRDKAEVCLWESQ---QSILEDEKLLISKTPVLPERAKEENGQOPR 1108  
Db 1109 -----ERFKLLQEFYEREGNTEDELEEEEDLPTASKLTPEDSPALTPPSPF 1156  
QY 1109 AANVCAGOSEELPP--KAVASKTENENLNOIGHOEKKTSSS 1147  
Db 1157 RDSVAGSSVSPSPVESVLCPTPNVYASVILRDYKQSSS 1197

ULT 15  
25763  
ID AAW25763 standard; Protein; 1219 AA.  
XX AC AAW25763;  
XX DT 03-DEC-1997 (first entry)  
XX DE Amino acid sequence of pCar/RI.  
XX KW Chimeric receptor; extracellular domain; seven transmembrane domain;  
KW intracellular cytoplasmic tail domain; metabotropic glutamate receptor;  
KW MGR; calcium receptor; CR; mGluR; G protein-coupled receptor;  
KW neurological disease.  
XX OS Chimeric - Homo sapiens.  
XX PN WO9705252-A2.  
XX PD 13-FEB-1997.  
XX PF 25-JUL-1996; 96WO-US12336.  
XX PR 26-JUL-1995; 95US-0001526.  
XX PA (NPSF-) NPS PHARM INC.  
PI Fuller FH, Hammerland LG, Krapcho KJ;  
WPI; 1997-145690/13.  
N-PSDB; AAT86166.  
XX Chimeric receptors comprising metabotropic glutamate receptor and  
PT calcium receptor - used for screening for neurologically active  
PT compounds  
XX Example 3; Fig 3; 177pp; English.  
CC This sequence represents a chimeric receptor. The chimeric receptor  
CC comprises an extracellular domain, a seven transmembrane domain, and  
CC an intracellular cytoplasmic tail domain, and a sequence of at least  
CC 6 contiguous amino acids is homologous to a sequence of a metabotropic  
CC glutamate receptor (MGR), and a sequence of at least 6 contiguous amino  
CC acids is homologous to a sequence of a calcium receptor (CR). The  
CC chimeric receptor may be used for screening for compounds that bind to  
CC or modulate the activity of MGR or CR and for determining the site-of-  
CC action of a CR active compound. The compounds can be used in the  
CC treatment of neurological diseases and disorders. They can also be  
CC used as diagnostic agents. Chimeric receptors such as this, allow the  
CC coupling of certain functional aspects of an MGR with certain functional  
CC aspects of a CR. They allow for more efficient high-throughput screening  
CC of compounds.

XX Sequence 1219 AA;  
Query Match 3.4%; Score 220; DB 18; Length 1219;  
Best Local Similarity 18.8%; Pred. No. 7.1e-07;  
Matches 174; Conservative 127; Mismatches 354; Indels 270; Gaps 34;  
QY 311 OKVDIDQSSDQWFGSTHCK---HLNNSSE-CMPIKGLGFLVLCAYECICKAG---FYHGV 363  
Db 475 EQVTFDEC---GDLVGNYSIIHNLSPEDGSIVFKEVGY---YVYAKKGERLFINEEK 527  
QY 364 LPVNNFRRGDPQH-----ISGSKDVSEE-----AYVCLPREG-----CPCAD 404  
Db 528 ILMSGFSREVFPNSCRDLAGTRKGIIEGTECTCFCEVCEPDGEYSDETDASACKNCPD 587  
QY 405 D-----SPCFVQEDKYLR-----LALISFOGLCMLLDFVSMVLVYHFRKAKSIRASG 451  
Db 588 DFWSNENHTSCPIPVRYLEWSDIESIIAIFSCILGILVTLFVTLFVLYRDTVPVKSSS 647  
QY 452 LILETILFGSLLLYFPVVILYFEPSTERCILLRWARILGFATVYGTVLKLRVLKV-- 509  
Db 648 RELCVIILAGIFLGVCFPTLIAKPTTSCYLORLLVGLSSAMCYSALVTNRIARILA 707  
QY 510 -----FLSRTAQRIPYMTGGRVMRLAVILLVFWFVLIGWTSVCQNLEKQIS 557  
Db 708 GSKKICKTRKPRFMSAWAQVIIASILISVQLTLVVTLII-----MEPMP 752  
QY 558 LIGQKTSDDLIFNCLIDRDWYMTAVA-EPLFLMGVYLCYAVRTVPSAPHEPRYMAVA 616  
Db 753 ILSYPSIKE--VYLICNTSNLGVAPVGYNGILLIMSCSYAFKTRNVNPNANFYIAFT 810  
QY 617 VHNELIISAIFHTIRFVLASRLQSDWMLMLYFAH"HLTVTITIGLLIPK-----FSHSSN 672  
Db 811 MYTCIILWAFVIYF-----GSNYKIITTCFAVLSVTVALGCMFTPKWVIIAKPER 864  
QY 673 NPRDDIATEAYEDELDMG-----RSGSYLNSSINSASWSEHSLDPEDIRDELKLYAQ 724  
Db 865 NVRSAFIT--SDVVRMHVGDGKLPSCRNTFLN-----894  
QY 725 LEIYKRRKMTNPHLOKRCR-----KGLGRSIRMRITEIPETVSRQSKEDKSGA 777  
Db 895 --IFRRKKPGAGNANSNGKSVSWSEPGGRQAPKQGVHVRWLSHVHVKTNETACNQ----- 946  
QY 778 DHGTAKGTALIRKNPPSSGNTGSKKEETLKNRVESLKKSHSTYDHRVDQTEESSSLPTE 837  
Db 947 -----TAVIK---PLTKSYOGSKSLTFSD-----968  
QY 838 SOEETTENSTLESLSGKKLTQKLKEDSEAESESTESVPLVCKSASAHNLSSEKKTGHPRTS 897  
Db 969 -----ASTKTLYNVEEDNTPSAHSPSPSPSMVVRHRRGPPVATTPPLPP 1013  
QY 898 MLQKSLSVIASAKEKTLGLAGKTQTAGVBEERTKSQKPLPKDKETNRNHSNDTETKQPA 957  
Db 1014 HL-----TAEETPLFADSVIPKGL-----PPPLPQQQ----- 1041  
QY 958 PONSNAPEPRKPKQSGIMKQORVNPNTTANSDL-----NPGTTQMKDNFDIGVCVCPW 1009  
Db 1042 PQQPPQPPQPPK--SLMDQLQGVVTFNGSGIPDFHAYLAGPGT-----GN----- 1087  
QY 1010 EYDITLPGFVPSE--SKVQKHVSIVASEMEKNPTFSLKCKSHHKPKAAEVCCSQSNOKRID 1067  
Db 1088 SLRSLYPPPPPPHQLQMLPLHLSTFOEESISPPGEDIDDD-----ERFK 1132  
QY 1068 KAEVCLWESQ---QSILEDEKLLISKTPVLPERAKEENGQOPRAANVCAGOSEELPP-- 1122  
Db 1133 LLOEFVYEREGNTEDELEEEEDLPTASKLTPEDSPALTPPSPFRDVSAGSSVSPSPVS 1192  
QY 1123 KAVASKTENENLNOIGHOEKKTSSS 1147  
Db 1193 ESVLCPTPNVYASVILRDYKQSSS 1217

Job time : 76.2905 secs

-----

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2003, 05:55:14 ; Search time 32.8197 Seconds  
(without alignments)  
3558.936 Million cell updates/sec

Title: US-09-775-181-2  
Perfect score: 6382  
Sequence: 1 MGAWAYPLLCLLQLAQLGLG.....LSANKIAGPRKEIWDSEFKV 1215  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	279.5	4.4	433	T21989	hypothetical prote
2	225	3.5	5327	T13564	microtubule-associ
3	220.5	3.5	1199	A41939	G protein-coupled
4	210	3.3	1171	A42916	metabotropic gluta
5	210	3.3	1180	JC2132	metabotropic gluta
6	204.5	3.2	6642	T29757	protein UNC-89 - C
7	201.5	3.2	1200	A46194	neurofilament prot
8	199.5	3.1	1212	JC2131	metabotropic gluta
9	194.5	3.0	3488	T34418	hypothetical prote
10	190.5	3.0	734	B42680	nucleolus-cytoplas
11	190	3.0	1390	T14004	trif protein - ali
12	188	2.9	1072	A86827	hypothetical prote
13	187	2.9	1046	T29776	hypothetical prote
14	186	2.9	522	C96608	hypothetical prote
15	183	2.9	852	T06310	hypothetical prote
16	183	2.9	1560	T42727	proliferation pote
17	181	2.8	1358	A29360	SIR4 protein - yea
18	179.5	2.8	1230	T22458	hypothetical prote
19	179	2.8	2004	F95133	immunoglobulin A1
20	178.5	2.8	1020	T29108	hypothetical prote
21	178.5	2.8	1963	B98002	IgA-specific metal
22	177.5	2.8	729	S68191	triadin - human
23	177.5	2.8	3924	T37431	ankyrin 2, neuro
24	177	2.8	532	T06029	hypothetical prote
25	177	2.8	667	A40713	cylicin I - bovine
26	177	2.8	1165	T16420	hypothetical prote
27	176	2.8	2094	S33124	tp protein - huma
28	175.5	2.7	1837	T41023	probable nuclear p
29	174.5	2.7	1131	A49393	activator 1 large

30	174	2.7	1020	1	QFHUH	neurofilament trip
31	174	2.7	1558	2	T29253	hypothetical prote
32	173	2.7	919	2	F81998	ribonuclease E (EC
33	173	2.7	990	2	I51618	nucleolar phosphop
34	172.5	2.7	706	2	A45990	junctional sarcopl
35	172.5	2.7	1079	2	I59362	calcium/polyvalent
36	172.5	2.7	1218	2	S71376	glutamate receptor
37	172	2.7	1310	2	T40135	probable involveme
38	171.5	2.7	971	2	T19431	hypothetical prote
39	171	2.7	1359	2	T34036	hypothetical prote
40	170.5	2.7	879	2	JC7160	metabotropic gluta
41	170	2.7	904	2	T24457	hypothetical prote
42	170	2.7	2218	2	B84683	hypothetical prote
43	169.5	2.7	598	2	B40713	cylicin I - human
44	169.5	2.7	916	2	A27864	neurofilament trip
45	169.5	2.7	919	2	F81225	ribonuclease E NNB

ALIGNMENTS

RESULT 1  
T21989  
hypothetical protein F39B2.8 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: T21989  
R:DOBSON, R.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19498  
A:Accession: T21989  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-433 <WIL>  
A:Cross-references: EMBL:Z92834; PIDN:CAB07388.1; GSPDB:GN00019; CESP:F39B2.8  
A:Experimental source: clone F39B2  
C:Genetics:  
A:Gene: CESP:F39B2.8  
A:Map position: 1  
A:Introns: 137/2; 249/2; 287/3; 338/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein F39B2.8

Query Match	4.4%	Score 279.5;	DB 2;	Length 433;
Best Local Similarity	28.7%;	Pred. No. 3.5e-08;		
Matches	102;	Conservative	57;	Mismatches 129; Indels 67; Gaps 13;
QY	456	ETILFGSLLLYPPVILY-----FEP-----STPRCIL---LRWARLLGFAIVYG	497	
Db	70	EHFLSTAVALKPRLTQLAHFFANHSKNFDPQLLDKSSFOQIPPFDFVSFARRTHFYGLQL	129	
QY	498	TVTCLKLHRLVKVFLSTAQRIPTYMTGGRYMRMLAVILLVFWFLIGMTSSVCQNLKQIS	557	
Db	130	NETGNHRLQYRVRKKAQHSV-VREQDMLKYLAAMLALTITGLMAWT-----	176	
QY	558	LIGQKTSDDLIFN---MCLIDRMDYTAFAEFLFLMGVLYCYAVRTVPSAFHEPRYM	613	
Db	177	---VSGWGTALRRATWPCQLMGQGHVHVGVELLFLLYAVRLCYKARN--SDWLERWQF	231	
QY	614	AVAVINELIISAFHTIRVLASRLQSDWMLMYFAHTHLTVTTVIGLLLIIPKFSHSSNN	673	
Db	232	TVAVCLEAVITLMANLIRYSIRNSGRADTLFVSVFVHLQTLTVSNVIVIVAPKPFYLSNGE	291	
QY	674	PRDDIATEAYEDELDMGRSG-----SYLNSSINSANSE-HSLDPEDIRDLKLL	721	
Db	292	PSRRSMTLG-----GHSGRAHPSLAKURDNLINGTIDFAEVPIDAMNPEDIRAEKRV	344	
QY	722	YAQLEIYKRKKMTNNPHLQKRCSSKGLGRS--IMRRITEIPETVSRQCSKEDK	774	
Db	345	YTQLRMVYKLNLYQDNPHISKRGGKKWSDKYTKATRRIS-IP-----SCSPQTK	393	

RESULT 2  
T13564

microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)  
N:Alternate names: hypothetical protein EG:49E4.1  
C:Species: Drosophila melanogaster  
C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13564  
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.  
submitted to the EMBL Data Library, April 1999  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: 217689  
A:Accession: T13564  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5327 <SPA>  
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1  
C:Genetics:

A:Cross-references: FlyBase:FBgn0025392  
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1  
A:Note: EG:49E4.1  
Superfamily: Drosophila 576k microtubule-associated protein homolog

Query Match 3.5%; Score 225; DB 2; Length 5327;  
Best Local Similarity 21.5%; Pred. No. 0.0011;  
Matches 128; Conservative 91; Mismatches 277; Indels 100; Gaps 22;

QY 668 SHSSNNPRDDIATEAYEDELMDGRSGSYLNSINSAMSEHSLDP-----EDIRDELKLYA 723  
DB 1716 AESAKDGADDLKELSPRESTTOSKEAGSIKDEKSPLEASEASRPASVAESVDEAKSKE 1775  
QY 724 QLEIYKRKMWTNPHLOKRSKGLGRSIMRRITEIPETVSRQCSKEDKGADHGTA 783  
DB 1776 E----SRRESVAEKSPLPSKEAS-----RPASVAESIKDEAKSKEERSRESVAE 1821  
QY 784 GTALIRK--NPPE-----SGTNGSKSEETLKNRPF-----SLKSHSYDHDVQDT 828  
DB 1822 KGPLPSKEASRPASVAESIKDEAKSKEERSRESVAEKSPLPSKEASRPASVAESIKDEA 1881  
QY 829 EBSLSLPTESQEBEETTENSTLES-----SGKLTQKLKEDSEAES--TESVPLVC 877  
DB 1882 EKSKE---ESRESVAEKSPLPSKEASRPASVAESIKDEAKSKEERSRESVAEKSPLPS 1938  
QY 878 KSA-----AHLNLSB---KKTGHPTSMLOKS-----LSVIASAKETLGLAGKTQTA 923  
DB 1939 KEASRPASVAESIKDEAKSKEERSRESVAEKSPLPSKEASRPASVAESIKDEAKSKEE 1998  
QY 924 GVEERTKSKOKPLKOKETNRHNSDNTETKDPAPONSNAEPRKPKQSGIMKQORVNP 983  
1999 SRRESVAEKSPLP-SKEASRPASVAES--IKDEAKSKEERSRESVAEKSPLPSKEASRP 2055  
984 TTANDLNPGTTOMKDNFDIGVCPWEVYDLTPGPVPSESKYQKHVSIIVASEMEKNPTFS 1043  
DB 2056 ASVAESIKDEAKSKE-----ESRESAAEKGPLPSKEASRPASVAESVKDEADKSKEE 2110  
QY 1044 LKEKSHHKPKAEV-CQOSNQRIDKAEVCLNESOGQSTLEDEKLLISKTPVLPRAKEE 1102  
DB 2111 RRESMAESGAQSIKGDQSLPEVSRP-----ESVAESVKD-----PVKSKEP 2154  
QY 1103 NGGQPRAAVCA--GQSEELP-----PKAVASKTENENLQIGHQKKTSSSEENV 1152  
DB 2155 SRRESVAGSVTADSRDQSPLESKGASRPESVDSVKDEAKQESRRRESKTESVIPKA 2214  
QY 1153 GSYNSNNFQPLSRAEYCPWEFFETPAQPNAGRSVALPASSALSANKIAGPRKEE 1208  
DB 2215 KDKSPKEVLQPV-SMTETIREDADQPMKPSQAES---RRESIAESIKASSPRDEK 2266

RESULT 3  
A41939  
G protein-coupled glutamate receptor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
C:Accession: A41939; S15362  
R:Houamed, K.M.; Kuijper, J.L.; Gilbert, T.L.; Haldeman, B.A.; O'Hara, P.J.; Mulvihill,  
Science 252, 1318-1321, 1991

A:Title: Cloning, expression, and gene structure of a G protein-coupled glutamate rec  
A:Reference number: A41939; MUID:92022526; PMID:1656524  
A:Accession: A41939  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1199 <HOU>  
A:Cross-references: GB:M61099; NID:g397806; PIDN:AAA19497.1; PID:g204460  
A:Experimental source: cerebellum  
A:Note: Sequence extracted from NCBI backbone (NCBIP:60785)  
R:Masu, M.; Tanabe, Y.; Tsuchida, K.; Shigemoto, R.; Nakanishi, S.  
Nature 349, 760-765, 1991  
A:Title: Sequence and expression of a metabotropic glutamate receptor.  
A:Reference number: S15362; MUID:91156047; PMID:1847995  
A:Accession: S15362  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1199 <MAS>  
A:Cross-references: EMBL:X57569; NID:g56646; PIDN:CAA40799.1; PID:g56647  
C:Keywords: G protein-coupled receptor; transmembrane protein

Query Match 3.5%; Score 220.5; DB 2; Length 1199;  
Best Local Similarity 18.6%; Pred. No. 0.00026;  
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;

QY 357 GFYHGVLPVNNRRRGGPDQHIHSGSTKDVSEE-----AYVCLPCREG- 398  
DB 498 GTWHEGLNIDDKYKIQ---MNSGVMVRSVCSEPLKGQIKVIRKGEVSCCWTACKENE 554  
QY 399 -----CPFC-----ADDSPCFVQEDKYLR-----LAISFGQLCMLDFFVSMVV 438  
DB 555 FVQDEFTCRACDLGNWPAELTGCEPIPVRYLEWSIDIESIIAFAESCLGILVTLFVLIF 614  
QY 439 YHFRAKSTRASGLLLETLFGSLLLPVVPVILFEPSTFRICILLRWALLGFATVYGT 498  
DB 615 VLYRDPVVPVSSRELCTIILAGIFLVGVCPFTLIAKPTTTCYQRLDLVGSSAMCYS 674  
QY 499 VTLKLRVLKV-----FLSRTAQRIPYMTGGRVARMNAVLLVVFVFLIGM 544  
DB 675 LVTKNRIARILAGSKKIKCTRKPRFMSAWAQVITASILISVQLTLVLTLLI----- 726  
QY 545 TSSVCONLEKQISLIGQGTSDHLINMCLIDRWDMYTAVA-EFLFLMGVLYCAVRIV 603  
DB 727 -----MEPPPIILSYPSIKE--VVLICNTSLNGVAPVGYNGLLIMSTCYAFAKTRV 777  
QY 604 PSAFHEPRYMAVAVENELIISAIFHTIREVLASRLQSDMMLYFAHTLTVTVTIGLL 663  
DB 778 PANFNEAKTIATMTTTCILWLAFPVTF-----GSNKYIITTCFANVSLSTVALGCMF 831  
QY 664 IPK----FSHSSNNPRDDIATEAYEDELDMG-----RSGSYLNSINSAMSEHSLDP 711  
DB 832 TPKMYIIIAKPERNVSAFTT---SDVVVMHVGDGKLPGRSNTFLN----- 874  
QY 712 EDIRDELKLYAQLAEIYKRKKMTNPHLOKRS-----KGLGRSIMRRITEIPT 764  
DB 875 -----IFRRKPKGAGNANSNGSVSWSEPGGRQAPKQGHVQWQRLSVHVKT 919  
QY 765 VSRQCSKEDKEGADHGTAQTALIRKNPESGNTCKSKEETLKNRVFSLKSHSTYD 824  
DB 920 NETAQN-----TAVIK---PLTKSYQSGSKSLTFSD----- 948  
QY 825 RQOTESSSLPTESQEBEETTENSTLESISGKKLTQKLKEDSEAESVPLVCKSASAHN 884  
DB 949 -----ASTKTLNVVEEDNTPSAHSPSPSPSPSVVHR 980  
QY 885 LSSEKKTGHPTSMLOKSLSVIASAKETLGLAGKTQTAGVVEERTKSKQPKPKDKETNRN 944  
DB 981 RGPVAVATTPLPPLH-----TAETPLFLADSVIPKGL-----PPPLPQQ----- 1021  
QY 945 HNSDNTETKDPAPONSNAEPRKPKQSGIMKQORVNPPTTANSDL-----NPGTQ 996  
DB 1022 -----POOPPPQPPQPPK--SLMDLQGVVTFNGSGIPDFHVLAVLAGPPT- 1065  
QY 997 MKDNFDIGVCPWEVYDLTPGPVPSE--SKVQKHVSIIVASEMEKNPTFSLKSHHKPKA 1054

Db 1066 -----GN-----SLRSLYPPPPPPHQLPLHLSTFOEESTSPGEDIDDD-----1108  
QY 1055 AEVCQSNOKRIDKAECVLWESOG---QSTLEDEKLLIKTPVLPERAKEENGQOPRAAN 1111  
Db 1109 -----ERFKLQEFYIERGNTTEDELEEDLPATSKLTPEDSPALTPSPFRDS 1159  
QY 1112 VCAGOSEELPP--KAVASKTENENLQIGHQEKTTSS 1147  
Db 1160 VASGSSVPSPVESVLCPTPNVYASVILRDYKQSSS 1197  
RESULT 4  
A42916  
metabotropic glutamate receptor mGluR5 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
C:Accession: A42916  
R:Abe, T.; Sugihara, H.; Nawa, H.; Shigemoto, R.; Mizuno, N.; Nakanishi, S.  
J. Biol. Chem. 267, 13361-13368, 1992  
A:Title: Molecular characterization of a novel metabotropic glutamate receptor mGluR5 cd  
A:Reference number: A42916; MUID:92317054; PMID:1320017  
A:Accession: A42916  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1171 <ABE>  
A:Cross-references: GB:D10891; NID:g220813; PIDN:BAA01711.1; PID:d1002186; PID:g220814  
A:Experimental source: brain  
A>Note: sequence extracted from NCBI backbone (NCBIN:107749, NCBI:P:107750)  
C:Keywords: G protein-coupled receptor; transmembrane protein  
Query Match 3.3%; Score 210; DB 2; Length 1171;  
Best Local Similarity 19.1%; Pred. No. 0.00097;  
Matches 179; Conservative 117; Mismatches 397; Indels 244; Gaps 35;  
QY 219 ETEWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDLGGDKSHFKWSPPYLECE 273  
Db 332 DVKWFDDYILKLPETNLRNP-----WFOEFWQHRRFCRLEGFAQENSKYNKT-----CN 381  
QY 274 NG-----SYKPGWLVTLSAIYGL--QPNLVPEFRGV-----MKVDINLQ 311  
Db 382 SSLLTTHVQDSKMGFVINAYSMAYGLHNMQMSLCPGYAGLCDAMKPIDGRKLDSLM 441  
QY 312 KVDIDQCSDDGFWFSGTHKCHLNNECMPIKGLGFLVYGAYECICKAGFYHPGVLPV-----366  
Db 442 KTNFTGVSGDMILFDENGSPGRYEIMNFKEMG---KDFDYINVGSWDNGELKMDDEV 498  
QY 367 ---NNFRRGPDQHI-SGSTKDVSEE---AYVCLPCREG-----CPFC-----A 403  
499 WSKNNIIRSVCSPECKGOIKVIRKEVSCCWTCTPCKENEVYVDEYTCACQLGSWPT 558  
404 DD-SPCFVQEDKYLR-----LAIISFOGLCMLLDFVSLVYVYHFRKAKSIRASGLILLE 456  
Db 559 DDLTGCDLIPVQVLRGDEPIAAVFAACGLLATLFTVTVIFIIYRDTPVVKSSRELKY 618  
QY 457 TILFGSLLYFPVVIYFFESTFRCLLRWARLLGFATYGVTVTLKHLRVLYK-----509  
Db 619 IILAGICLGLYCLFLCIAKPKYICVQLRIGLISPAWSYALVTNRIARILAGSKK 678  
QY 510 -----FLSRTAQRIPYTGGRVMEMLAVILLVWFLLIGHTSSVCNLEKOISL-----558  
Db 679 ICTKPRFWSACQ-----LVIAFILI-----CIQLGIIVAFIME 714  
QY 559 ---IGQGKTSDDLIFNMCLIDRDYMTAVA-EFLFLWGLYLCYAVRTVPSAFHEPRYMA 614  
Db 715 PPDIMHDYPSIRVYILICNTNLGVVTPICYNGLLLISCTFTFAFKTRNVPANFNKAYIA 774  
QY 615 VAVHNELIISAIFHTIRFVLASRLQSDMMLYFAHTLTVTIGLLIPK-----FSHS 670  
Db 775 FTWYTTCTIWLAFVPIYF-----GSNKIITMCFVSLSATVALGCMFVPKVIILAKP 828  
QY 671 SNPNRDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLLDPDIRDELKK-----720

Db 829 ERNRSFAFTTSTVVRMHVGDGKSSSAASRSSSLNVLNWKRRSGSETLSSNGKSVTWQAQNE 888  
QY 721 -----LYAQLEIYKRRKMTNNPHLQK---KRCSSKGLGRSIMRRITEIPETVSRQCS 770  
Db 889 KSTRGHLWQRLSVHINKENKENTQAVIKPPFKSTENRG-----PCAAAGGGS 936  
QY 771 KEDKEGADHGTAKGTALIRKNPPESSGNTGKSEETLKNRVFSLKSHSYTHVHRDQTEE 830  
Db 937 GPGVAGA--GNAGCTATGGPEPPDAGPKA-----LYDVAEA 970  
QY 831 SSSLPTESQEEETENSTLESLSGKKLTOKLEDSEABTESVPLVCKSASAHNLSSEKK 890  
Db 971 EESFPAARPRSPSPITSLHLAG-----SAGRTDDDDAPSLSHSETAARSSSSQGS 1020  
QY 891 TGHPRTSMLOKLSLVIASAKERTGLAG-----KQTACAVEER 928  
Db 1021 LMEQISSVTRTANISELSNMMLSSTAATPGPGTGPICSSYLIPKELQPTMTTFAE-- 1078  
QY 929 TKSQPLPKDKETNRNHSNDTETKDPAPQNSNPAEE-PRKPQSGIMKQORVNPTTAN 987  
Db 1079 ---IQPLPAIEVT---GGAQGATGVSQAETPTGAESAPGRKPDLEELVALTPPSP---- 1127  
QY 988 SDLNPCTTQMKONFDIGEVCWPHEVYDLTGPVPYSESK 1024  
Db 1128 -----FRDSVDSGSTTPNSPVSESALCIPSSPK 1155  
RESULT 5  
JC2132  
metabotropic glutamate receptor 5 A - human  
C:Species: Homo sapiens (man)  
C:Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 05-Jan-1996  
C:Accession: JC2132  
R:Minakami, R.; Katsuki, F.; Yamamoto, T.; Nakamura, K.; Sugiyama, H.  
Biochem. Biophys. Res. Commun. 199, 1136-1143, 1994  
A:Title: Molecular cloning and the functional expression of two isoforms of human met  
A:Reference number: JC2131; MUID:94197696; PMID:7908515  
A:Accession: JC2132  
A:Molecule type: mRNA  
A:Residues: 1-1180 <MIN>  
C:Comment: this protein is coupled to guanine nucleotide binding proteins.  
C:Keywords: glycoprotein; neurotransmitter; receptor; transmembrane protein  
F:580-604/Domain: transmembrane #status predicted <TM1>  
F:617-637/Domain: transmembrane #status predicted <TM2>  
F:644-664/Domain: transmembrane #status predicted <TM3>  
F:694-714/Domain: transmembrane #status predicted <TM4>  
F:738-759/Domain: transmembrane #status predicted <TM5>  
F:773-794/Domain: transmembrane #status predicted <TM6>  
F:803-827/Domain: transmembrane #status predicted <TM7>  
Query Match 3.3%; Score 210; DB 2; Length 1180;  
Best Local Similarity 19.1%; Pred. No. 0.00098;  
Matches 179; Conservative 120; Mismatches 400; Indels 240; Gaps 34;  
QY 219 ETEWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDLGGDKSHFKWSPPYLECE 273  
Db 333 DVKWFDDYILKLPETNLRNP-----WFOEFWQHRRFCRLEGFPQENSKYNKT-----CN 382  
QY 274 NG-----SYKPGWLVTLSAIYGL--QPNLVPEFRGV-----MKVDINLQ 311  
Db 383 SSLLTTHVQDSKMGFVINAYSMAYGLHNMQMSLCPGYAGLCDAMKPIDGRKLDSLM 442  
QY 312 KVDIDQCSDDGFWFSGTHKCHLNNECMPIKGLGFLVYGAYECICKAGFYHPGVLPVNFR 371  
Db 443 KTNFTCVSGDTLFDENGSPGRYEIMNFKEMG-----KDFDYINVGSWDNGEL 492  
QY 372 RGPDQHI---SGSTKDVSEE-----AYVCLPCREG-----CPFC 402  
Db 493 KMDDDEWSSKSNIRSVCSPECKGOIKVIRKEVSCCWTCTPCKENEVYVDEYTCAC 552  
QY 403 -----ADD-SPCFVQEDKYLR-----LAIISFOGLCMLLDFVSLVYVYHFRKAKSIRA 449  
Db 553 QLGSWPTDDLITGCDLIPVQVLRGDEPIAAVFAACGLLATLFTVTVIFIIYRDTPVWKS 612









Db 561 SSEEKKPKSKA-----TP-KPQAGKANGVPASQ-----NGRAGKESEE 599

RESULT 11

T14004

trfa protein - slime mold (Dictyostellium discoideum)

C:Species: Dictyostellium discoideum

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-May-2000

C:Accession: T14004

R:Saito, J.; Adachi, H.; Sutoh, K.

J. Biol. Chem. 273, 24654-24659, 1998

A:Title: Dictyostellium TRFA homologous to yeast Ssn6 is required for normal growth and e

A:Reference number: Z17852; MUID:98406112; PMID:9733762

A:Accession: T14004

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1390 <SAI>

A:Cross-references: EMBL:AB009080; NID:d1228566; PID:d1034109; PIDN:BAA33143.1

A:Gene: trfa

Introns: 333/3; 364/3; 637/1

Query Match 3.0%; Score 190; DB 2; Length 1390;

Best Local Similarity 21.1%; Pred. No. 0.016;

Matches 111; Conservative 81; Mismatches 228; Indels 106; Gaps 20;

QY 664 IPKFSSNNPRDDIATEAYEDELDMGRSGSYLN-----SSINSAWSHSLDPEDIRDE 717

Db 901 ISRFHONRTSSPENNNNNNNNNNNNNINNICORALSPQSSQHKORREILIDE 960

QY 718 LKKLAQLEIYKRRKMITNPHLOKKRCSK-----GLGRSMRRITEIPETVSRQCSKE 772

Db 961 -----ESDINERSK--TRSPSIVKAEAEKRETVIVDKERSPTPIITE-----KP 1002

QY 773 DKEGADHGTAQTALIRK-----NPRESSONTGSKB---ETLKNRVFSLUKSHSYDHYVR 825

Db 1003 DEKQVEKVTDKESLVKDYDKENEKESPSSSSSSSEIEKETEKEKEKEKEVEKE 1062

QY 826 DOTESESSLPTESQ--EEETENTSTLESGLKLTOKLKDSEASTESVPLVCKSASAHN 884

Db 1063 KEIENDKEKEKEKEKDYENKSKSEKPV-----EKEST-----TTTTND 1106

QY 885 LSEKKTGHPTSMLOKSUSVIAAKERTGLAGTQTAGVEERTKSKPLPKDKETNRN 944

Db 1107 EDEGELSEPTTTTKDDSKLPT-DEKLLSSVPTTTA-VEQSRDETKELMDTKEDSE 1164

QY 945 HNSDNTETKDPAPQNSPAEPRPKQSGIMKQORVNPPTTANSDLNPCTTQMKDNFDIG 1004

Db 1165 KEKSGSTTTTAAASESVKPIDBEKK-----SPTTTTTTTTTTTT 1203

QY 1005 EYCPVEVYDLTPGPVPESEKQKHVSIVASEMEKNPTFSLKEKSHHKPKAAEVCQOSNOK 1064

Db 1204 -----VEPHDKESKNDTTTTTTTTTTTSAKSPNSPTRSDEVEPHQDASOE 1254

QY 1065 RIDKAEVLWESQGOSEILDEKLLISKTPVLPRAKEENGGOVPRANVCAGOSEELPPKA 1124

Db 1255 EINKRK-----LEDD---ITSTP--SKRLKPDSS--TPSSATTASTPSEQ--PES 1294

QY 1125 VASKTENENLNOIGHQEKKTSTSEENVRGYSSNNNFQOPLTSRAE 1170

Db 1295 PUKK-ENPVGETLSPBIKDKSSSSSSSSSSSTNTGSSSTNSAK 1339

RESULT 12

A86827

hypothetical protein yqfG [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C>Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C:Accession: A86827

R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: A86827

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1072 <STO>

A:Cross-references: GB:AE005176; PID:gl2724625; PIDN:AAK05715.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: yqfG

Query Match 2.9%; Score 188; DB 2; Length 1072;

Best Local Similarity 18.9%; Pred. No. 0.014;

Matches 146; Conservative 109; Mismatches 264; Indels 254; Gaps 32;

QY 618 HNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTLTV-----TVTIGILL 663

Db 176 NNNLTSLGSGNITTEIMAKAFQSQITITIEFADESITVDSAAFSGSSVQSITLIGVTL 235

QY 664 I-----PKFSSNNPRDDIATEAYEDELDMGRSGSYLNSSINS----- 702

Db 236 ADDVFNKTSPLFGQLSDLPTDEIRT-----ISVNSSGLFDKSWINSDDSSQSTSSTENA 289

QY 703 ----AWSEHSLDPEDIRDELKKL---YAOLEYIKRKKMITNPHLOKKRCSKKGGLGRSTM 755

Db 290 DTPVSSNSNVQVDVASSSESTQDANSASLYPISEASVTDN-----TL 333

QY 756 RRITEIPETVSRQCSKEDEGA-----DHGTAQTALIRKKNPPSSGNTGK-SKEETLK 808

Db 334 NSISLDSISSQSTENSQSGASSTAEISYDSENSNSLSSNQINSNSEKDSQSSGLG 393

QY 809 NRVSFLSKS-HSTYDHRVDQTEESSL-----PTE-----SQEETENTSTLES 851

Db 394 SSMSSNEEHESSNSNI-NETNNSSEITNLPSPNPTESNSVSDQTSSEASTNSSSI-S 451

QY 852 LSGKKLTOKLKDSEAESTE-----SVPLVCKSASAHNL----- 885

Db 452 LSPSNISSDSESATNSDFNSVAEAVANNLSAVNNSSSVLSTSTADNLGINQSGSD 511

QY 886 -----SSEKKTGH-----PRTSMLOKLSL----- 904

Db 512 NUTKDSSEISTSGAFSSNOTSSEASTNSNISLSIPNSISSTSVLESTSSNSFNSVAE 571

QY 905 ----VIASAKERTGLAGTQTA-----GVEERTKSQKPL-----PKDKETN 942

Db 572 VANNSLASVNNSSSVLSSTSTADNLEINQFGSDNLTKDSSEISTSGAFSSNOTSSEAS 631

QY 943 RNHSNSDN-----TETKDPAPQNSPAEPRPKQSGIMKQORVN-----PTTAN 987

Db 632 SNMSGINSPLSLSLTNSSEATNQSNSEATKYVDNNSSTHSSNLSNGSGNSDSDSD 691

QY 988 SFLNPCTTQMKDNFDIGVC---PWEVYDLTPGP--VPSESKVQKHVSIVASEMEKNPTF 1042

Db 692 SUDSSNLSNLSNLETNOTISSKPEVNINSENPKVSSNSVQENST--DHENSTNPKS 749

QY 1043 SLKEKSHHKPKAAEVCQOSNQKRIDKAEVCLWESQGOSEILDEKLLISKTPVLPRAKEE 1102

Db 750 SIS-----SPITSTSSQOK-----ESQSNLLNTTEGI---NNPITFNSSSE 789

QY 1103 NGQOPRAANVCA-----GQSEELPPKAVASKTENENLNOIGHQEKKTSSSEENVRGYN 1156

Db 790 NS-----AASILTSYNNNSSESSETGCLYISNEAORDNGSEISHSLPSSNSNNV----- 840

QY 1157 SSNNFQOPLTSRAEVCPEFETPAQPNAGRSVALPASSALS-ANKTAGPRKEE 1208

Db 841 SSIQSOAILLES-----SKSTNKRKSSLSLSTNSTSHPQNEED 876

RESULT 13

T29776

hypothetical protein C50F2.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T29776



Query Match 2.9%; Score 183; DB 2; Length 852;  
Best Local Similarity 19.6%; Pred. No. 0.02;  
Matches 121; Conservative 92; Mismatches 211; Indels 192; Gaps 23;

QY 680 TEAYEDELDMGRSG-----SYLNSINSINSAWSEHSLDPEDIRDELKKLYAQLEIYKRKMI 734  
Db 293 TDAPKDE--SGKSGVSGVNGVAQQNDSDVTDMSKKDDTGAKDEPOOL----- 337

QY 735 TNNPH---LQKKRCCKGLGRIMRRITETVSRQCSKED--KEGADHGTAAGTALIR 789  
Db 338 -DNPRNTDLNNTTEKPDVEHQIEEKENESSVKQADLSKDSDIKEETEPEALLDSKQVL 396

QY 790 KNPPESSNGTGSKEETLKNRVFSKKSHSTYDHRDQTESSSLPTESQEEETENSTL 849  
Db 397 TSPVPDVSSVTAATSSNEKNKSVQILPSKTSGD-----ETANVSSPSMAELPE--- 445

QY 850 ESLSGKKLTQKLKEDSEAESESTESVPLVCKSASAHNLSEKKTGHPRTSMLOKSLSVIASA 909  
Db 446 QSVPRKTAQNRKKKSESTEVEKPSA-----SIATEEVSEEPNTSEPOVT--KSGKKVASS 498

QY 910 KEKTGLAGKTOTAGVEERTKOKPLPKDKETNRNHSNDNTETK-----DPAPQNSNPAE 965  
Db 499 -----SKTKPVPPSPKSTSETKVAKQSEKKVVGSDNAQESTKPKKE 539

QY 966 EPRKPKQSGIMKQORVNPPTANSDLNPGTOMKDNFDIGEVCPWEYDITPGVPVSESKV 1025  
Db 540 EKKKPGRGKAIDEESLH-----TSSGDN-----EKPAVSSGKLASKSKK 578

QY 1026 QKHVSIVASEMEKNPTFSLKEK-SHHKPKAA----- 1055  
Db 579 EAKQTV-----BESPNSTNRKRSLGQKGASGESLVGSRIKVWMPMDQAYYKGVVESYDA 633

QY 1056 -----EVCQOSNOK--RIDKAEVCLWE----- 1075  
Db 634 AKKKHLVIYDDGQDEILYLKQKWSPLDESELQDEEADQGTQGEDASTVSGGAGSSKA 693

QY 1076 -----SQGOSILEDEKLLISKTPVLPERAKEENGQOPRAANVCAGQSEELPPKAYA---- 1126  
Db 694 KATPASKSKTSQDDK-TASKSKDSKEASREBEASSEES-----EEEEPPKTVGKSGS 746

QY 1127 --SKTENENLIGHQEKKTSSSEENVRGSYNSNNFQOPLTSRAEVCWFEFETPAQPN 1184  
Db 747 SRSKXDISSVKSG-KSKASSKKKEEPSKATTSSKSKGPFVKS-----VPAKSKT 795

QY 1185 GRSVALPASSALSANK 1200  
Db 796 GKGAAGSGASTPASK 811

Match completed: February 15, 2003, 08:32:03  
Run time : 57.8197 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2003, 04:22:33 ; Search time 18.7541 Seconds  
(without alignments)  
2687.074 Million cell updates/sec

Title: US-09-775-181-2

Perfect score: 6382

Sequence: 1 MGAMAYPLLCLLAQLGLG.....LSANKIAGPRKEIWDSEFKV 1215

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	220.5	3.5	1199	1 MGR1_RAT	P23385 rattus norv
2	211	3.3	1203	1 MGR5_RAT	P31424 rattus norv
3	201.5	3.2	700	1 TRDN_CANFA	P82179 canis famil
4	199	3.1	704	1 NP14_RAT	P41777 rattus norv
5	197.5	3.1	1212	1 MGR5_HUMAN	P41594 homo sapien
6	193	3.0	1194	1 MGR1_HUMAN	Q13255 homo sapien
7	183	2.9	976	1 MGR_DROME	P91685 drosophila
8	181	2.8	1358	1 SIR4_YEAST	P11978 saccharomyc
9	178	2.8	1637	1 MRSP_STAAU	P80544 staphylococ
10	177.5	2.8	728	1 TRDN_HUMAN	Q13061 homo sapien
11	177.5	2.8	3924	1 ANK2_HUMAN	Q01484 homo sapien
12	177	2.8	667	1 CYLL_BOVIN	P35662 bos taurus
13	176	2.8	2349	1 TPR_HUMAN	P12270 homo sapien
14	175.5	2.7	1079	1 CASR_MOUSE	Q9Gy96 mus musculu
15	174.5	2.7	1131	1 AC15_MOUSE	P35601 mus musculu
16	174	2.7	1020	1 NFH_HUMAN	P12036 homo sapien
17	172.5	2.7	633	1 MLH_TETTH	P40631 tetrahymena
18	172.5	2.7	705	1 TRDN_RABIT	Q28820 tryctolagus
19	172.5	2.7	1079	1 CASR_RAT	P48442 rattus norv
20	172	2.7	1310	1 YB35_SCHPO	Q14340 schizosacch
21	171	2.7	1359	1 ATRX_CAEEL	Q9U7e0 caenorhabdi
22	169.5	2.7	598	1 CYLL_HUMAN	P35663 homo sapien
23	169.5	2.7	915	1 NFM_HUMAN	P07197 homo sapien
24	169	2.6	999	1 MGR1_CAEEL	Q09630 caenorhabdi
25	168.5	2.6	1233	1 YFI6_YEAST	P43597 saccharomyc
26	168.5	2.6	1257	1 RBB1_HUMAN	P29374 homo sapien
27	168.5	2.6	1849	1 IGA4_HAEIN	P45386 haemophilus
28	168	2.6	1411	1 TCOF_HUMAN	Q13428 homo sapien
29	167.5	2.6	816	1 YG3A_YEAST	P53278 saccharomyc
30	167	2.6	836	1 NOT3_YEAST	P06102 saccharomyc
31	166	2.6	2492	1 ATRX_HUMAN	P46100 homo sapien
32	165.5	2.6	952	1 IF41_YEAST	P39935 saccharomyc
33	165	2.6	1275	1 TRP_DROME	P19334 drosophila

RESULT 1	
MGR1_RAT	
ID	STANDARD; PRT: 1199 AA.
AC	P23385;
DT	01-NOV-1991 (Rel. 20, Created)
DT	01-NOV-1991 (Rel. 20, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Metabotropic glutamate receptor 1 precursor (mGluR1).
GN	GRM1 OR GPRC1A OR MGLUR1.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Sprague-Dawley; TISSUE=Brain;
RX	MEDLINE=91156047; PubMed=1847995;
RA	Masu M., Tanabe Y., Tsuchida K., Shigemoto R., Nakanishi S.;
RT	"Sequence and expression of a metabotropic glutamate receptor.";
RL	Nature 349:760-765(1991).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Brain;
RX	MEDLINE=92022526; PubMed=1656524;
RA	Houamed K.M., Kuijper J.L., Gilbert T.L., Haldeman B.A., O'Hara P.J.,
RA	Mulvihill E.R., Almers W., Hagen F.S.;
RT	"Cloning, expression, and gene structure of a G protein-coupled
RT	glutamate receptor from rat brain.";
RL	Science 252:1318-1321(1991).
RN	[3]
RP	ALTERNATIVE SPLICING (ISOFORM 1B).
RC	TISSUE=Brain;
RX	MEDLINE=92110002; PubMed=1309649;
RA	Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;
RT	"A family of metabotropic glutamate receptors.";
RL	Neuron 8:169-179(1992).
RN	[4]
RP	ALTERNATIVE SPLICING (ISOFORM 1C).
RC	TISSUE=Brain;
RX	MEDLINE=93066232; PubMed=1438218;
RA	Pin J.-P., Weeber C., Prezeau L., Bockaert J., Heinemann S.F.;
RT	"Alternative splicing generates metabotropic glutamate receptors
RT	inducing different patterns of calcium release in Xenopus oocytes.";
RL	Proc. Natl. Acad. Sci. U.S.A. 89:10331-10335(1992).
CC	!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
CC	MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-
CC	CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL
CC	ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM POTENTIATION IN
CC	THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.
CC	!- SUBCELLULAR LOCATION: Integral membrane protein.
CC	!- ALTERNATIVE PRODUCTS: 3 ISOFORMS: 1A (SHOWN HERE) 1B AND 1C;
CC	ARE PRODUCED BY ALTERNATIVE SPLICING. 1B AND 1C ARE C-TERMINALLY
CC	TRUNCATED FORMS OF 1A.
CC	!- TISSUE SPECIFICITY: IS PREDOMINANTLY EXPRESSED IN CEREBELLAR
CC	PURKINJE CELLS, CA2-CA3 PYRAMIDAL CELLS OF THE HIPPOCAMPUS, AND
CC	MITRAL AND TUFTED CELLS OF THE OLFACTORY BULB.

P18583 homo sapien  
P06180 xenopus lae  
Q61687 mus musculu  
Q14832 homo sapien  
O84419 chlamydia t  
P46012 caenorhabdi  
P70486 rattus norv  
P31422 rattus norv  
Q02952 homo sapien  
O43493 homo sapien  
P98193 rattus norv  
P30414 homo sapien

34 165 2.6 2426 1 SON\_HUMAN  
35 164.5 2.6 589 1 HIBN\_XENLA  
36 164 2.6 2476 1 ATRX\_MOUSE  
37 163.5 2.6 877 1 MGR3\_HUMAN  
38 163.5 2.6 1770 1 PMPC\_CHLFR  
39 162.5 2.5 952 1 YK15\_CAEEL  
40 162 2.5 527 1 ATRX\_RAT  
41 162 2.5 879 1 MGR3\_RAT  
42 162 2.5 1781 1 AKAC\_HUMAN  
43 161 2.5 480 1 TGN2\_HUMAN  
44 161 2.5 489 1 DMP1\_RAT  
45 161 2.5 1462 1 NKCR\_HUMAN









CC A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES.  
CC -!- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND  
CC DEPHOSPHORYLATION ON CK-II AND PKC SITES. NOPPI40 IS ONE OF THE  
CC MOST PHOSPHORYLATED PROTEINS IN THE CELL.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M94287; AAA41718.1; -.  
CC EMBL; M94288; AAA41719.1; -.  
KW Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding.  
FT DOMAIN 84 570  
FT ACIDIC SERINE CLUSTER.  
FT REPEAT 84 95  
FT ACIDIC SERINE CLUSTER 1.  
FT REPEAT 127 138  
FT ACIDIC SERINE CLUSTER 2.  
FT REPEAT 170 181  
FT ACIDIC SERINE CLUSTER 3.  
FT REPEAT 231 242  
FT ACIDIC SERINE CLUSTER 4.  
FT REPEAT 274 285  
FT ACIDIC SERINE CLUSTER 5.  
FT REPEAT 335 346  
FT ACIDIC SERINE CLUSTER 6.  
FT REPEAT 373 384  
FT ACIDIC SERINE CLUSTER 7.  
FT REPEAT 434 445  
FT ACIDIC SERINE CLUSTER 8.  
FT REPEAT 479 490  
FT ACIDIC SERINE CLUSTER 9.  
FT REPEAT 524 535  
FT ACIDIC SERINE CLUSTER 10.  
FT REPEAT 559 570  
FT ACIDIC SERINE CLUSTER 11.  
FT MOD\_RES 567 567  
FT PHOSPHORYLATION (BY CK2).  
FT VARIANT 150 150  
FT MISSING (IN NOPPI40B).  
SQ SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;

Query Match 3.1%; Score 199; DB 1; Length 704;  
Best Local Similarity 21.3%; Pred. No. 0.0036;  
Matches 126; Conservative 92; Mismatches 240; Indels 134; Gaps 24;

QY 665 PKFSSHNNP-----RDDIATEAYED-----ELDMGR-----SGSYLN 697  
DB 64 PKVKLQNGPVAKKAKKETSSDSDSEEDKQVPTQKAAAPAKRASLPQHAGKAA 123  
QY 698 SSINSAWSHSLDPEDIRDELKLLAQLAIIYRKMKMTNNPHLQKRCCKGLGRSIRMR 757  
DB 124 KASESSSSSESEEBEKKKKVQQAQKAVPQAKAVRPPP---KKAESSESSESSS- 179  
QY 758 ITEIPETVSRQCKDEKADGHTAGTALIRKNPPSESSNTGKSKKEETLKNRVFSLKKS 817  
DB 180 -DEAPVT-----QKPKAAATAAKAPTQAOTKAPKAPGPPAKQPKAANGKAGSSSS 230  
818 HSTYDVRDQTEBSSSLPTSESQEEETTENSTLESLSGKK-LTQKLKEDSEAESTESVPLV 876  
231 -----SSSSSDSDSEEEKAAAPLTKTAPKQVQVAKAPVKVTAFTQ----- 272  
QY 877 CKSASAHNLSE-----KKTGHPRTSMLOKLSUSVITASKETKTLGLAGKTOTAGVEE 927  
DB 273 -KSSSDSDSESEEBEQKPMKKAGPYSSVPPSVSL-----SKKSVG----- 315  
QY 928 RTKSKPLPKDETNRNHSNNTETKDPAPONSNAPEPRPKPSGIMK-QQRVNPPTA 986  
DB 316 ---AQSPKKAQAQTQPADSASDSEED-----SSSEEEKTKTAKTVVSKTTPAKPAPVK 367  
QY 987 NSDLNPGTTQMDKDNFDIGVCPWEVYDLTPGPVPSKVKQKHVSIVASDMKKNPTFSLKE 1046  
DB 368 KAE---SSSDSDSDSEDEAPAKPVSAFKSPLSKRPVPTPKPAKAVATPKQPA----- 419  
QY 1047 KSHHKPAEAVCOQSNQKIDKAEVCLWESQSQSTLEDEKLISK-TPVLPPEAKEENG 1105  
DB 420 GSGQKPKQRKADSSSESESSSSSE---EATKKSVTTPKARVTAAPSLPAKQAPRAGG 476  
QY 1106 QPRAANVCAGQSEE--LPKAVASK-----TENENLQIGHQEKKTSSSENVRGSGN 1156  
DB 477 DSSSDSSSESEEEKTPPKPAKKAAGAAPPKTPPVKKAASSSSSSSSE-----D 530

QY 1157 SSNFFQQLTSRAEYCPWFEFETPAQPNAGRSVALPASSALSANKIAGPRKEE 1208  
DB 531 SSEEKKKPKSKA-----TP-RPQAGKANGVPASQ-----NGRAGKESEE 569

RESULT 5  
MGR5\_HUMAN  
ID MGR5\_HUMAN STANDARD; PRT; 1212 AA.  
AC P41594;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-OCT-1996 (Rel. 34, last sequence update)  
DE 15-JUN-2002 (Rel. 41, last annotation update)  
DE Metabotropic glutamate receptor 5 precursor (mglur5).  
GN GRM5 OR GPRC1E OR MGLUR5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=Brain;  
RX MEDLINE=94197696; PubMed=7908515;  
RA Minakami R., Katsuki F., Yamamoto T., Nakamura K., Sugiyama H.;  
RT "Molecular cloning and the functional expression of two isoforms of  
human metabotropic glutamate receptor subtype 5";  
RL Biochem. Biophys. Res. Commun. 199:1136-1143(1994).  
RN [2]  
RP REVISIONS.  
RA Katsuki F.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 860-952 FROM N.A.  
RX TISSUE=Brain;  
RX MEDLINE=93343913; PubMed=7688218;  
RA Minakami R., Katsuki F., Sugiyama H.;  
RT "A variant of metabotropic glutamate receptor subtype 5: an  
evolutionally conserved insertion with no termination codon";  
RL Biochem. Biophys. Res. Commun. 194:622-627(1993).  
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS  
CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-  
CC CALCIUM SECOND MESSENGER SYSTEM AND GENERATES A CALCIUM-ACTIVATED  
CC CHLORIDE CURRENT.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 5A (SHOWN HERE) AND 5B; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER BY AN INSERTION OF  
CC 32 RESIDUES.  
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.  
CC STRONGEST, TO MGLUR1.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D28538; BAA05891.1; -.  
CC EMBL; D28539; BAA05892.1; -.  
CC EMBL; S64316; AAD13954.1; -.  
CC Genew; HGNC:4597; GRM5.  
CC MIM; 604102; -.  
CC InterPro; IPR001828; ANF\_receptor.  
CC InterPro; IPR000337; GPCR\_Mgr.  
CC Pfam; PF00003; 7tm\_3; 1.  
CC Pfam; PF01094; ANF\_receptor; 1.  
CC PRINTS; PR00248; GPCRMR.  
CC PROSITE; PS00979; G\_PROTEIN\_RECEP\_F3\_1; 1.  
CC PROSITE; PS00980; G\_PROTEIN\_RECEP\_F3\_2; 1.  
CC PROSITE; PS00981; G\_PROTEIN\_RECEP\_F3\_3; 1.  
CC PROSITE; PS00981; G\_PROTEIN\_RECEP\_F3\_3; 1.  
CC PROSITE; PS00981; G\_PROTEIN\_RECEP\_F3\_3; 1.  
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
KW Multigene family; Alternative splicing.











DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
 KW Antibiotic resistance; Glycoprotein; Cell wall; Peptidoglycan-anchor;  
 Repeat; Signal.  
 FT SIGNAL 1 48  
 FT CHAIN 49 1601  
 FT PROPEP 1602 1637  
 FT DOMAIN 1301 1582  
 FT SITE 1598 1602  
 FT MOD\_RES 1601 1601  
 FT SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;  
 Query Match 2.8%; Score 178; DB 1; Length 1637;  
 Best Local Similarity 20.8%; Pred. No. 0.12;  
 Matches 99; Conservative 71; Mismatches 197; Indels 108; Gaps 17;  
 QY 763 EIVSQCSEKEDGADHGHTAKTALIRNPPSSNGTCKSKEETLKNRVSLKSHSYD 822  
 47 EAAENNTTQKDDSDASKVGNV---QTIEQSSAN---SNESDIPEQV-----D 90  
 823 HVRDQTEESS-----LPTESOEETSTNLSLGGKLTOKL-----EDSEA 867  
 91 VTKDTEQASTEKANTTEQASTEKADTEQATTEAPKAEQDKVTEEPAPKAEETDK 150  
 868 ESTESVPLVCKSASAHNLSEKKTGHPRTSMLOKSLSVIASAKET-----LGLAGKT 920  
 151 AUTEAP---KAEDTKATEA---PKTEETDKATTEAPAAEETSKAATEAPKAEET 203  
 QY 921 QTAGVEERTKSQ-----KPLPKDKEINRHNS- -DNTETKDPAPQNSNPAEPRKPQ 971  
 204 SKAATEAPKAEETKATEAPKTEETDKVTEEPAPKAEETSKAATEAPKAEETNKVE 263  
 QY 972 KSGIMKQORVNPPTANSDLNPGTQKDNFDIGVCPWEVDLTGPPVPSKVKHYSI 1031  
 264 TEAPAAETNKAAETEPAVEDTNAKS-----SNAOPSEFRTQVVD 308  
 QY 1032 VASEMEKNPTFLSKKESH-HKPKAAEVCQSNQKRIDKAEVCLWESQGSILEDEKLLIS 1090  
 309 VAKDLYKKSEVTEAKAEIEKVLPLKDISLNE-----EIKKIALS 349  
 QY 1091 KTPVLPERAKENGOPRAANVACQSELPKPKAVASKTENENLQIGHQEKKTSSSEN 1150  
 350 E--VLKETANKEN-AOPRAF-----RSVSSNARTNNVSNATLRAAAQDTVT 395  
 QY 1151 VRGSYSSNNFOQPLTSRAEVCPEW-----FETPAQPNAGRSVALPASSALSANK 1200  
 396 KKGTFNTHAGDIHHTYKEEPPNGLTAFNTNPNTPGKALEYNDKIDFNK 450  
 RESULT 10  
 ID TRDN\_HUMAN STANDARD; PRT; 728 AA.  
 AC Q13061;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE Triadin.  
 GN TRDN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX MEDLINE=96061957; PubMed=7588753;  
 RA Taske N.L., Eyre H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,  
 RA Foster P.S.;  
 RT "Molecular cloning of the cDNA encoding human skeletal muscle triadin  
 RT and its localisation to chromosome 6q22-6q23.";  
 RL Eur. J. Biochem. 233:258-265(1995).  
 CC -1- FUNCTION: MAY BE INVOLVED IN ANCHORING CALSEQUESTRAIN TO THE  
 CC JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL  
 CC COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic  
 CC reticulum.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; U18985; AAA75315.1; -  
 CC Genew; HGNC:12261; TRDN.  
 CC MIM; 603283; -  
 KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein.  
 FT INIT\_MET 0 0  
 FT DOMAIN 1 46  
 FT TRANSMEM 47 67  
 FT POTENTIAL 68 728  
 FT LUMENAL  
 FT CARBOHYD 74 74  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 646 646  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 728 AA; 81423 MW; C1C53BBE1B2A0815 CRC64;  
 Query Match 2.8%; Score 177.5; DB 1; Length 728;  
 Best Local Similarity 20.3%; Pred. No. 0.047;  
 Matches 139; Conservative 95; Mismatches 291; Indels 161; Gaps 24;  
 QY 642 WMLMIFYAHTHTVVTI-LGILLIPKFSHS-----NNPRDIATEAYEDELDMGRSGSY 695  
 Db 49 WLLVIALIITWSAVAIVMFDLVYKFNFSASSTAKTIGSDPL-KLVRDAMEETDWTGYFFS 107  
 QY 696 LNSINSANSSEHSIDPEDIRDE-----LKKLYAQLLEYKRR-----KMTNNPH 739  
 Db 108 LLSDIISSEDEDDGDGDDTDKGEIDEPPLRR-----KEIHKDKTEKQEKPKKIQTKVTH 163  
 QY 740 LQKRCSCKGLGRSIRMTIPEITVSQCSK-EDKEADHGHTAGTALIRKNPPESSGN 798  
 Db 164 KEKEKGEK-----VREKEKPEKATHEKIEKKEPEKTKVAKEQKAKTAEKSEK 216  
 QY 799 TKG-----SKEETLNRPVSLKSHSYDHRDQTEESS----- 833  
 Db 217 TKVEYKGGKQKVKQTAAKVKEVQKTPSKPKKEKDEKAAVSKHEQKQDQYAFCRYMIDIF 276  
 QY 834 -----LPTESOEETSTNLSLGGKLTOKLDESEASTESVPLVC 877  
 Db 277 VHGLKPCQSPAIPLPTEQASRPTPASPALEKEKEKKAEEKVTSETKKKEKEDIK 336  
 QY 878 KSAHNLSEKKTGHPRTSMLOKSLSVIASAKETTLGLAGKTQTAGVEERTKSOKPL-- 935  
 Db 337 KSEKETAIIDVEKEPGRASETKQGTVKIAQAQAAK-----KDEKEDSKTKKPAEV 388  
 QY 936 --PKDKETNRHNSDNTETKDPAPQNSNPAEPRKPKQSGIMKQORVNPPTANSDLN-- 991  
 Db 389 EQPKGKQKKEKHEV-----PAKSPKKEHSVPSDKQVKAETKAEETGAV 435  
 QY 992 -----PGTOMKDNFDIGVCPWEVDLTGPPVPSKVKQ-KHVSIVASEMEKNPTFSL 1044  
 Db 436 SSKAVPGKKEKTKTVQEIREEKSGKTSILKDKPEIKGKEKVPASLKEKEPETKK 495  
 QY 1045 KEK-----SHHKPKAAEVCQSNQKRIDKAEVCLWESQGSILEDEKLLISKTPVL-PER 1098  
 Db 496 DEKMSKAGKEVKPKP---PQLOGKKEKEPEPQIKKEAKPAI--SEKVOIHKQDIVKPEK 549  
 QY 1099 AKE-----ENGQPPRAANVACQSEBELP-----PKAVASKTEN 1131  
 Db 550 TVSHGKPEEKVLQKVAVTIEKTAKPKTKKAHREPRPPSIKTDKPKTKPGTSEVTES 609  
 QY 1132 ENLQIGHQEKKTSSSENVGYSNNSNNFOPLTSRAEVCPEWFEFETPAQPNAGRS---V 1188  
 Db 610 -----GKKKTEISEKSEKADMKHLREEKYSTRKESLQHLHNVTKAEKPARVSKDVE 661  
 QY 1189 ALPAS-SALSANKIAGPRKEIWDSEF 1213  
 :||| | :| :| :|



Db 662 DVPASKAKEGTEDVSTKQKSPISF 687

RESULT 11

ANK2\_HUMAN

ID ANK2\_HUMAN STANDARD; PRT; 3924 AA.

AC Q01484; Q01485;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).

GN ANK2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).

RC TISSUE=Brain stem;

RX MEDLINE=91302466; PubMed=1830053;

RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;

"Isolation and characterization of cDNAs encoding human brain ankyrins reveal a family of alternatively spliced genes.";

J. Cell Biol. 114:241-253(1991).

RN [2]

RN REVISIONS.

RA Carpenter S.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Brain stem;

RX MEDLINE=94075409; PubMed=8253844;

RA Chan W., Kordeli E., Bennett V.;

"440-kD ankyrinB: structure of the major developmentally regulated domain and selective localization in unmyelinated axons.";

J. Cell Biol. 123:1463-1473(1993).

RN [4]

RP SEQUENCE OF 463-495 FROM N.A.

RX MEDLINE=9200921; PubMed=1833308;

RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,

Lux S.E., Ward D.C., Forget B.G.;

"Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.";

Genomics 10:858-866(1991).

CC -!- FUNCTION: Attach integral membrane proteins to cytoskeletal elements. Also bind to cytoskeletal proteins.

CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL CELLS THROUGHOUT THE BRAIN.

CC -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE AND FUNCTION (POTENTIAL).

CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.

CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announcement/or\\_send\\_an\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announcement/or_send_an_email_to_license@isb-sib.ch)).

CC EMBL; X56957; CAA40278.1; -

CC EMBL; X56958; CAA40279.2; -

CC EMBL; Z26634; CAA42644.1; -

CC EMBL; M37123; AAG62828.1; -

CC PIR; S14533; S14533.

CC PIR; A39643; A39643.

CC PIR; B39643; B39643.

CC PIR; S14569; S14569.

CC HSP; P42771; IDC2.

DR Genew; HGNC:493; ANK2.

DR MIM; 106410; -

DR InterPro; IPR002110; ANK.

DR InterPro; IPR000488; Death.

DR InterPro; IPR000906; ZU5.

DR Pfam; PF00023; ank; 24.

DR Pfam; PF00531; death; 1.

DR Pfam; PF00791; ZU5; 1.

DR PRINTS; PR01415; ANKYRIN.

DR SMART; SM00248; ANK; 21.

DR SMART; SM00005; DEATH; 1.

DR SMART; SM00218; ZU5; 1.

DR PROSITE; PS50088; ANK\_REPEAT; 20.

DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.

DR PROSITE; PS50017; DEATH\_DOMAIN; 1.

KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;

KW Phosphorylation.

FT REPEAT 63 92 ANK 1.

FT REPEAT 96 125 ANK 2.

FT REPEAT 129 158 ANK 3.

FT REPEAT 162 191 ANK 4.

FT REPEAT 193 220 ANK 5.

FT REPEAT 232 261 ANK 6.

FT REPEAT 265 294 ANK 7.

FT REPEAT 298 327 ANK 8.

FT REPEAT 331 360 ANK 9.

FT REPEAT 364 393 ANK 10.

FT REPEAT 397 426 ANK 11.

FT REPEAT 430 459 ANK 12.

FT REPEAT 463 492 ANK 13.

FT REPEAT 496 525 ANK 14.

FT REPEAT 529 558 ANK 15.

FT REPEAT 562 591 ANK 16.

FT REPEAT 595 624 ANK 17.

FT REPEAT 628 657 ANK 18.

FT REPEAT 661 690 ANK 19.

FT REPEAT 694 723 ANK 20.

FT REPEAT 727 756 ANK 21.

FT REPEAT 760 789 ANK 22.

FT REPEAT 793 822 ANK 23.

FT DOMAIN 1773 1950 REPEAT-RICH REGION.

FT REPEAT 1773 1784 REPEAT A.

FT REPEAT 1785 1796 REPEAT A.

FT REPEAT 1797 1808 REPEAT A.

FT REPEAT 1809 1820 REPEAT A.

FT REPEAT 1821 1832 REPEAT A.

FT REPEAT 1833 1844 REPEAT A.

FT REPEAT 1845 1856 REPEAT A.

FT REPEAT 1857 1867 REPEAT A (APPROXIMATE).

FT REPEAT 1868 1879 REPEAT A.

FT REPEAT 1880 1891 REPEAT A.

FT REPEAT 1892 1902 REPEAT A (APPROXIMATE).

FT REPEAT 1903 1914 REPEAT A.

FT REPEAT 1915 1926 REPEAT A.

FT REPEAT 1927 1938 REPEAT A.

FT REPEAT 1939 1950 REPEAT A.

FT DOMAIN 3536 3620 DEATH.

FT VARSPLIC 1039 1039 (IN ISOFORM 2).

FT VARSPLIC 1444 3528 MISSING (IN ISOFORM 2 AND ISOFORM 3).

FT CONFLICT 475 476 GQ -> PE (IN REF. 4).

FT CONFLICT 971 971 I -> S (IN REF. 1).

FT CONFLICT 3581 3582 QY -> HA (IN REF. 1).

FT CONFLICT 3586 3586 I -> Y (IN REF. 1).

SQ SEQUENCE 3924 AA; 52AC496C428E29D2 CRC64;

Query Match 2.8%; Score 177.5; DB 1; Length 3924;

Best Local Similarity 21.3%; Pred. No. 0.4;

Matches 123; Conservative 85; Mismatches 230; Indels 139; Gaps 26;

QY 691 RGSYL--NSSINSWSEHSL-----DPEDIRDELKKLYAQLEI----- 727

DB 1534 RSGTCTRDESSVQSSRSRGLVEEHWIVSDEIEEARQK--APLEITEYPCVEVRIDKE 1591

```
QY 728 -----YKFKMTNNP-----HLQK-----KRCCKGLGRSIRMRITEIPE 763
DB 1592 IKKVKEDSTGLVNYLTDLNTCVPLPKEQLTVQDQAGKKCEALAVGRSSEKEGDIPP 1651
QY 764 TVSRQCKEDKEGADHGATAGTALIRKNPPSESGNTGKSEETLKNRV--FSLKKSHTSY 821
DB 1652 DEQSTQKHPSL--GIKPP--VRRKLKEQ----KQKEGLQASAEKALKGSSEE 1702
QY 822 DHVRDQTESSSLPTESQEEETENSTLESQK-KLTKLKEDSEASTESVPLVCKS- 879
DB 1703 SLGEDPGLAEPPLTVKATSPLEETPIGSIKDKVALQKRVED-EQKGRSKLPPIRVKG 1761
QY 880 -----ASAHNLISSEKTHGHPRTSMLOKLSLVITASAKK---TLGLAGTKTQ-- 921
DB 1762 EDVPKKTHRPHAPASPLSKSRHAPGSPKTERHUSTIUSSSAKTERHPVPSPSSKTEKH 1821
QY 922 -TAGVEERTKSQPLPKDKETNRHNSDNTETKDPAPQNSNPAEPRPKQSG-IMKQQ 979
DB 1822 SPVSPSAKTERHSPASSSSKTEKHSPVSPSTKTERHSPVSSSTKTERHPVPSGKTDKRP 1881
QY 980 RVNP---TTANSLNPGTQMKDNFDIGEVCPEVVDLTPG-----PVPSSKVKQHV 1029
DB 1882 PVPSPGRTKHPVSPGRTKRP-----LPVSPSGRTDKHPVSTAGKTEKHL 1928
QY 1030 SIVAS-EMKKNPTFSLKEKSHHKPKAAEVCQ-----OSNQKRIDKAEVCLWE-----1075
DB 1929 PVPSPGRTKHPVSPSTKTERTEETMSVRELKATQSGODP-SKHKTGLFEHKSQKQ 1987
QY 1076 --SQGOSILEDEKLLSKTPVLPERAKEENGOPRAANCAGQSEELPPKAVASKTENEN 1133
DB 1988 PQEKGVRVEKE-----KGPILTQREAO-----KTENQTIKRGORLPVTGTAESKRGVR 2036
QY 1134 LNOIGHQEKTSSEENVRGSSNNNFQPLTSRAE 1170
DB 2037 VSSIGVKVEDAGGKEKV-----LSHKIPEVQSVPE 2068

RESULT 12
CYLL_BOVIN
ID CYLL_BOVIN STANDARD; PRT; 667 AA.
AC P35662;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cylicin I (Multiple-band polypeptide I).
DI CYLI OR CYL.
OC Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
RT "Molecular characterization of mammalian cylicin, a basic protein of
the sperm head cytoskeleton."
RL J. Cell Biol. 122:1043-1052(1993).
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -!- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
CC EMBL; 222779; CAA80456.1;
DR PIR; S35913; S35913.
DR PIR; A40713; A40713.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 287 569 9 APPROXIMATE TANDEM REPEATS.
FT REPEAT 287 305 1.
FT REPEAT 306 337 2.
FT REPEAT 338 368 3.
FT REPEAT 369 405 4.
FT REPEAT 406 442 5.
FT REPEAT 443 475 6.
FT REPEAT 476 516 7.
FT REPEAT 517 547 8.
FT REPEAT 548 569 9.
FT DOMAIN 617 667
SQ SEQUENCE 667 AA; 74817 MW; CBF66EA462243D91 CRC64;

Query Match 2.8%; Score 177; DB 1; Length 667;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 103; Conservative 89; Mismatches 180; Indels 138; Gaps 19;

QY 666 KFSHNNPRDDIATDAYEDELDMGRSGSYLNSINSANSEHSLDPEDIR-DELKKLYAQ 724
DB 228 KYSKSKNNSDAVSETCSKNSSNVGLM-VHLGES-----DAESMEFDWMLKNYSQ 276
QY 725 LEIYKRRKMITNNPHLOKKRCSK---KGLGRSIRMRITEIPETVSRQCSKEDKEGADHGT 781
DB 277 -----NNSKKPTTKDAKDAKGGSD-----AESVDSKDAKKKATKDT 317
QY 782 AKGTALIRKNPESGNT---GKSKEETLKNRVFSLK---SHSTYDHYVRDQTESSSLP 835
DB 318 KKGAKDTESTDAESGSDAKDAKKGKSKKKDAKDAKDAESGSDAKDAKDAESGSDAKDAK 377
QY 836 TESQEE-----ETTENTLESLSGKLTQKLKEDS-----EAESTESVP 874
DB 378 KDSKKDNKKDAKDAESTDAESGSDAKDAKKGKSKKKDAKDAKDAESTDA-- 435
QY 875 LVCKSASAHNLSEKTHGHPRTSMIQ-KSLSVIASAKEKTLGLACKTQTAGVEERTKSOK 933
DB 436 ----ESGDSNAKKDKSKKKDKDAKDAKDAVSTDAESEGDGAKSK----KDSKKDKK 488
QY 934 PLPKOKETNRHNSDNTETKDPAPQNSNPAEPRPKQSGIMKQORVNPPTTANSNLNG 993
DB 489 DLKDDQKPKAMKSESTETESDWSKVKRSDKDTKTA---KKATESGAESD---- 541
QY 994 TTOMKDNFDIGEVCPEVVDLTPGPVPSKVKQKHVSIIVASEMEKNPTFSLKSHHKPK 1053
DB 542 -----VSSKRYLKKTEMFKSSDAE-----SEESLFRP- 568
QY 1054 AAEVQCSNQKRIDKAEVCLWESQGSQSILEDKLLISTKTPVLPERAKEENGQ----- 1106
DB 569 -----GSKRVDESATSTDSKDAVEPRKGRIMPSRRTTFKKGKIGTGRVPPSPS 621
QY 1107 -----PRAANVCAGQSEELPPK 1123
DB 622 RPLPPECEILPSRVKRLCRCOMPMPPPK 651

RESULT 13
TPR_HUMAN
ID TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoprotein TPR.
GN TPR
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```



Chang W., Tu C., Chen T.-H., Komuves L., Oda Y., Pratt S.A., Miller S., Shoback D.; "Expression and signal transduction of calcium-sensing receptors in cartilage and bone."; Endocrinology 140:5883-5893(1999). [4]

SEQUENCE OF 507-582 FROM N.A. (ISOFORM A). STRAIN=NMRI; TISSUE=Brain; Hildenbrand J., Ammon H.P.T., Wahl M.A.; Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases. [5]

SEQUENCE OF 562-814 FROM N.A. TISSUE=Kidney; Moawad T.I., Riccardi D.; Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases. [6]

SEQUENCE OF 646-799 FROM N.A., AND TISSUE SPECIFICITY. MEDLINE=97231187; PubMed=9076582; Quarles L.D., Hartle J.E. II, Siddhanti S.R., Guo R., Hinson T.K.; "A distinct cation-sensing mechanism in MC3T3-E1 osteoblasts functionally related to the calcium receptor."; J. Bone Miner. Res. 12:393-402(1997).

-1- FUNCTION: SENSES CHANGES IN THE EXTRACELLULAR CONCENTRATION OF CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.

-1- SUBCELLULAR LOCATION: Integral membrane protein.

-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE PRODUCED BY ALTERNATIVE SPLICING.

-1- TISSUE SPECIFICITY: EPIDERMIS, KIDNEY AND CARTILAGE.

-1- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF110178; AAD28371.1; -  
EMBL; AF110179; AAD28372.1; -  
EMBL; AF128642; AAD40638.1; -  
EMBL; AF068900; AAC19388.1; -  
EMBL; AB027140; BAA77688.1; -  
EMBL; AF002015; AAC53252.1; -  
EMBL; AF159565; AAF00193.1; -  
MGD; MGI:1351351; Gprc2a.  
InterPro; IPR001828; ANF\_receptor.  
InterPro; IPR000337; GPCR\_Mgr.  
Pfam; PF00003; 7tm.3; 1.  
Pfam; PF01094; ANF\_receptor; 1.  
PRINTS; PR00248; GPCRNGR.  
PROSITE; PS00979; G\_PROTEIN\_RECP\_F3\_1; 1.  
PROSITE; PS00980; G\_PROTEIN\_RECP\_F3\_2; 1.  
PROSITE; PS00981; G\_PROTEIN\_RECP\_F3\_3; 1.  
PROSITE; PS0259; G\_PROTEIN\_RECP\_F3\_4; 1.  
G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
Alternative splicing.

SIGNAL 1 19  
FT CHAIN 20 1079 POTENTIAL.  
FT DOMAIN 20 612 EXTRACELLULAR CALCIUM-SENSING RECEPTOR.  
FT TRANSMEM 613 635 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 636 649 I (POTENTIAL).  
FT TRANSMEM 650 670 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 671 680 II (POTENTIAL).  
FT TRANSMEM 682 700 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 701 724 III (POTENTIAL).  
FT TRANSMEM 725 745 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 746 769 IV (POTENTIAL).  
FT TRANSMEM 770 792 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 793 805 V (POTENTIAL).  
FT TRANSMEM 806 828 VI (POTENTIAL).

```
DB 949 TLOPOOOOPOOPRCKQKQVIFGSGTWTFSLSFDEPKQKNAHMRNMQNSLEAQSNDTL 1008
QY 766 SR-----QCSKEDKEGADHGTAKGTALIRKNPPSSGNTGKSKEETLKNRVFSLKKS 817
DB 1009 NRHQALLPQCAEADSEMTIQTGTGQWPVGHQPEI-----ESPDEMPALVSTSRK 1062

RESULT 15
AC15_MOUSE
ID AC15_MOUSE STANDARD; PRT; 1131 AA.
AC P35601;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Activator 1 140 kDa subunit (Replication factor C large subunit) (A1
DE 140 kDa subunit) (Rf-C 140 kDa subunit) (Activator 1 large subunit)
DE (A1-P145) (Differentiation specific element binding protein)
DE (ISRE-binding protein).
DE RFC1 OR RECC1 OR IBF-1.
OS Mus musculus (Mouse).
NCBI_TaxID=10090;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
[1]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=94089669; PubMed=8265586;
RA Burbello P.D., Utani A., Pan Z., Yamada Y.;
RT "Cloning of the large subunit of activator 1 (replication factor C)
RT reveals homology with bacterial DNA ligases.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11543-11547(1993).
[2]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94158835; PubMed=8114700;
RA Luckow B., Bunz F., Stillman B., Lichter P., Schuetz G.;
RT "Cloning, expression, and chromosomal localization of the
RT 140-kilodalton subunit of replication factor C from mice and
RT humans.";
RL Mol. Cell. Biol. 14:1626-1634(1994).
[3]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss;
RX MEDLINE=95388065; PubMed=7659092;
RA McGehee Habener J.F.;
RT "Differentiation-specific element binding protein (DSEB) binds to a
RT defined element in the promoter of the angiotensinogen gene required
RT for the irreversible induction of gene expression during
RT differentiation of 3T3-L1 adipoblasts to adipocytes.";
RL Mol. Endocrinol. 9:487-501(1995).
[4]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RA Haque S.J.;
RL Submitted (FEB-1994) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE OF 1-565 FROM N.A.
RC STRAIN=LAF1;
RA Lössie A.C., Haugen B.H., Wood W.M., Camper S.A., Gordon D.F.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
[6]
RN SEQUENCE OF 354-528 FROM N.A.
RX MEDLINE=90229765; PubMed=1691767;
RA Haque S.J., Kumar A., Fischer T., Rutherford M.N., Williams B.R.;
RT "Evaluation of inter- and intramolecular primary structure homologies
RT of interferons by a Monte Carlo method.";
RL J. Interferon Res. 10:31-31(1990).
CC -!- FUNCTION: THE ELONGATION OF PRIMED DNA TEMPLATES BY DNA POLYMERASE
CC DELTA AND EPSILON REQUIRES THE ACTION OF THE ACCESSORY PROTEINS
CC PCNA AND ACTIVATOR 1. THE 140 SUBUNIT BINDS TO THE PRIMER-
CC TEMPLATE JUNCTION.
CC -!- SUBUNIT: HETEROPTAMER OF SUBUNITS OF 140/145, 40, 38, 37, AND
CC 36.5 kDa THAT FORMS A COMPLEX WITH PCNA IN THE PRESENCE OF ATP.
```

```
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE ACTIVATOR 1 140 KDA SUBUNIT FAMILY.
CC -!- SIMILARITY: CONTAINS 1 BRCT DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U01222; AA21643.1; -
CC EMBL; X72711; CAA51260.1; -
CC EMBL; U36441; AAA79698.1; -
CC EMBL; U07157; AAC52140.1; -
CC EMBL; U15037; AAB60452.1; -
CC MGD; MGI:97891; Reccl.
CC InterPro; IPR003959; AAA_ATPase_cent.
CC InterPro; IPR001357; BRCT.
CC InterPro; IPR000862; RfCdomain.
CC Pfam; PF00004; AAA; 1
CC Pfam; PF00533; BRCT; 1.
CC SMART; SM00292; BRCT; 1.
CC PROSITE; PS0172; BRCT; 1.
KW DNA replication; ATP-binding; Transcription regulation; DNA-binding;
KW Activator; Nuclear protein; Zinc-finger.
FT DOMAIN 399 477 BRCT.
FT NP_BIND 635 642 ATP (BY SIMILARITY).
FT ZN_FING 734 751 C2HC-TYPE (POTENTIAL).
FT DOMAIN 1104 1108 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 354 528 INTERFERON-STIMULATED-RESPONSE-ELEMENT
FT BINDING REGION.
FT CONFLICT 66 66 Y -> N (IN REF. 3).
FT CONFLICT 187 187 E -> EPDFCLSLFFGIQ (IN REF. 4).
FT CONFLICT 254 254 V -> A (IN REF. 5).
FT CONFLICT 559 559 N -> S (IN REF. 4).
FT CONFLICT 614 614 MISSING (IN REF. 3 AND 4).
FT CONFLICT 945 945 S -> N (IN REF. 1).
FT CONFLICT 1071 1071 T -> A (IN REF. 3).
FT CONFLICT 1104 1104 K -> Q (IN REF. 4).
SQ SEQUENCE 1131 AA; 125984 MW; A6F4F970A7F9EE94 CRC64;

Query Match 2.7%; Score 174.5; DB 1; Length 1131;
Best Local Similarity 23.6%; Pred. No. 0.12;
Matches 108; Conservative 76; Mismatches 157; Indels 117; Gaps 23;

QY 761 IPEVTSROCKEDKEGADHGTAKGTALIRKNPPSSGNTGKSKEETLKNRVFSLKSHST 820
DB 17 VNETHV-----KNEKTKASEGTGKGGKGVKAKVNNSG-----KEDASKPKQHS-KKKRII 65
QY 821 YDHVRDQTEESSSLPTESQEEETTE-----NSTLESLSGK--KLTOKLKEDSEAEATES 872
DB 66 YD-----SDSESETVQVKNNAKKKSEKLSYKPGKVSQKDPVTYVSETDED 112
QY 873 VPLVCKSAHNLSEKKTG-----HPRTSMLOKLSLSVIAKAKETGLAGKTQTA---- 923
DB 113 DDFVCCKAA-----SKSKENGVSNTSYLGTSNVKKNENV-KTKNKPSPKILTPTSVLDY 167
QY 924 -GVEERTKSKQPL--PKDKETNRNHSND-NTETKDPAPQNSNPAAEPKPKQKSGIMKQ 979
DB 168 FGTESVQRSGKKMWTSKRKSSQNTEDSRNLDEATAKQLQDDEDAELERQ-----LHEDE 222
QY 980 RVNPTTANSLNPGTQMKDNFDICEVCPWEVYDLTPGPVPSESKVQKHVSIVASEMEKN 1039
DB 223 EFARTLALDEEPKTKARKDSEGE-----ESFSSVQDDLS--KAEKQKS 266
QY 1040 PT----FSLKEKS-----HKPKAAE-----VCOOSNQKRI-----DKAEVLWESQOGSI 1081
DB 267 PNKALFLTSTARTKTSYPAKHGKRASEDAKQCKSAHREACSSPKASAKLALMKAKESS 326
QY 1082 LEDEKLLSKTPTVLPRAKEENGQPRANVCAGOSEELPPKRAVASKTENENLNIQHGE 1141
```

Db 327 YNETELLAAR-----RKESATEPKG-----EKTTPKKTQVSPTKRESVSPEDSEK 371  
Qy 1142 KKTSSSEENVRGSYNSSNNFOOP-----LTSRAEVC 1172  
Db 372 KRTNYQ-----AYRSYLNREGPKALGSKEIPKGAENC 403

Search completed: February 15, 2003, 08:27:12  
Job time : 41.7541 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2003, 04:37:08 ; Search time 67.6488 Seconds  
(without alignments)  
3700.688 Million cell updates/sec

Title: US-09-775-181-2  
Perfect score: 6382  
Sequence: 1 MGAWAYPLLCLLQAQLGLG.....LSANKIAPRKEIWDSEFKV 1215

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 205047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2811	44.0	4 Q9ULT3	Q9ult3 homo sapien
2	1532.5	24.0	6 Q9GKT3	Q9gkt3 macaca fasc
3	571.5	9.0	5 Q9VR40	Q9vr40 drosophila
4	408.5	6.4	5 Q45500	O45500 caenorhabdi
5	307	4.8	5 Q9VKA3	Q9vka3 drosophila
6	281	4.4	5 Q9WLP3	Q9wlp3 drosophila
7	254	4.0	5 Q9VL133	Q9vl133 drosophila
8	243.5	3.8	5 Q9BML6	Q9bml6 drosophila
9	238.5	3.7	5 Q9WIP4	Q9wip4 drosophila
10	237.5	3.7	13 Q98UC4	Q98uc4 gallus gall
11	232	3.6	5 Q9BML5	Q9bml5 drosophila
12	227.5	3.6	5 Q9VPS7	Q9vps7 drosophila
13	225	3.5	5 Q9SY55	Q9sy55 drosophila
14	225	3.5	5 Q96891	O96891 drosophila
15	222.5	3.5	11 Q9EPV6	Q9epv6 mus musculus
16	215.5	3.4	13 Q98UC5	Q98uc5 gallus gall

17	209	3.3	1160	11 Q9ESC8	Q9esc8 mus musculus
18	205.5	3.2	3111	5 Q9VH10	Q9vh10 drosophila
19	205	3.2	1183	5 Q9W3D1	Q9w3d1 drosophila
20	204.5	3.2	6632	5 Q17362	Q17362 caenorhabdi
21	204.5	3.2	6632	5 Q01761	Q01761 caenorhabdi
22	201.5	3.2	1200	5 Q25388	Q25388 loligo peal
23	200	3.1	2081	10 Q9LH98	Q9lh98 arabidopsis
24	198.5	3.1	1812	5 Q9VZD9	Q9vzd9 drosophila
25	198	3.1	2083	5 Q9N435	Q9n435 caenorhabdi
26	197	3.1	679	5 Q95XW8	Q95xw8 caenorhabdi
27	195	3.1	1338	11 Q9JIX8	Q9jix8 mus musculus
28	194.5	3.0	3484	5 P91257	P91257 caenorhabdi
29	194	3.0	1156	13 Q98UC6	Q98uc6 gallus gall
30	193	3.0	1695	5 Q9BKL2	Q9bkl2 hydra atten
31	193	3.0	1481	5 Q9VFR6	Q9vfr6 drosophila
32	190.5	3.0	602	5 Q8SY16	Q8sy16 drosophila
33	190	3.0	1390	5 Q77033	Q77033 dictyostell
34	189	3.0	808	4 Q9UK88	Q9uk88 homo sapien
35	188	2.9	1029	5 P91176	P91176 caenorhabdi
36	188	2.9	1072	16 Q9CF64	Q9cf64 lactococcus
37	187.5	2.9	734	5 Q8T1A0	Q8t1a0 dictyosteli
38	186	2.9	522	10 Q9FXB5	Q9fxb5 arabidopsis
39	185	2.9	3201	5 Q9W0U2	Q9w0u2 drosophila
40	183	2.9	852	10 Q9S255	Q9sz55 arabidopsis
41	182.5	2.9	1163	4 Q9UHB7	Q9uhb7 homo sapien
42	181.5	2.8	1698	2 Q9LC00	Q9lc00 staphylococ
43	181	2.8	2151	5 Q9VLL3	Q9vll3 drosophila
44	181	2.8	2151	5 Q9NG79	Q9ng79 trichomonas
45	181	2.8	2151	5 Q9VPL6	Q9vpl6 drosophila

## ALIGNMENTS

## RESULT 1

Q9ULT3 PRELIMINARY; PRT; 544 AA.  
AC Q9ULT3;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TRENBLrel. 13, Last annotation update)  
DE KIAA1136 protein (Fragment).  
GN KIAA1136.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=20039618; PubMed=10574461;  
RA Hirotsawa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;  
RT "Characterization of cDNA clones selected by the Genemark analysis  
RL from size-fractionated cDNA libraries from human brain.";  
RL DNA Res. 6:329-336(1999).  
DR EMBL; AB032962; BAA86450.1; .  
FT NON\_TER 1  
SQ SEQUENCE 544 AA; 60195 MW; DA1175E7926041F7 CRC64;

Query Match 44.0%; Score 2811; DB 4; Length 544;  
Best Local Similarity 99.8%; Pred. No. 4.8e-172;  
Matches 543; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	672	NNPRDIIATEAYEDELDMGRSGSYLNSSINSWSEHSLDPEDIRDELKLYLAQLEIYKRR	731
DB	1	NNPRDIIATEAYEDELDMGRSGSYLNSSINSWSEHSLDPEDIRDELKLYLAQLEIYKRR	60
QY	732	KMITNPHLOKRCCKGGLGRSIMRRITPEITVSRQCSKEDKEGADHGTAGTALIRKN	791
DB	61	KMITNPHLOKRCCKGGLGRSIMRRITPEITVSRQCSKEDKEGADHGTAGTALIRKN	120
QY	792	PPSSGNTGCKSEETLKNRVFSLKKSHSYDHRQOETSSSLPTSESQEEETSTLES	851

```
Db 121 PPSSGNTGKSEETIKNRVSLKSKSHSTYDHRVQDTESSSLPTSESQBEETTSTLES 180
QY 852 LSGKKLTOKLKEDSAAESTESVPLVCKSASAHNLSEKKTGHPTSMLOKSLSVTASAKE 911
Db 181 LSGKKLTOKLKEDSAAESTESVPLVCKSASAHNLSEKKTGHPTSMLOKSLSVTASAKE 240
QY 912 KTLGLAGKTQTAGVBEERTKSKQPLPKDKETNRHNSONTETKDPAPQNSNPAEPRKPQ 971
Db 241 KTLGLAGKTQTAGVBEERTKSKQPLPKDKETNRHNSONTETKDPAPQNSNPAEPRKPQ 300
QY 972 KSGIMQQRVNPTTANSDLNPGTQTKMKNFDIGVCPWEVYDLTPGVPSPSEKVKQHVSI 1031
Db 301 KSGIMQQRVNPTTANSDLNPGTQTKMKNFDIGVCPWEVYDLTPGVPSPSEKVKQHVSI 360
QY 1032 VASEMEKNPTFSLKEKSHHKPAAEVCQSNOKRDKAEVCLWESQOGSILEDEKLLISK 1091
Db 361 VASEMEKNPTFSLKEKSHHKPAAEVCQSNOKRDKAEVCLWESQOGSILEDEKLLISK 420
Db 1092 TPVLPERAKEENGOPRAANVCAGOSEELPPKAVASKTENENLNGIGHQEKKTSSSEENV 1151
Db 421 TPVLPERAKEENGOPRAANVCAGOSEELPPKAVASKTENENLNGIGHQEKKTSSSEENV 480
QY 1152 RGSYSSNNFQOPLTSRAEVCWPWEETPAQPNAGRSVALPASSALSANKIAGPRKEEIID 1211
Db 481 RGSYSSNNFQOPLTSRAEVCWPWEETPAQPNAGRSVALPASSALSANKIAGPRKEEIID 540
QY 1212 SFKV 1215
Db 541 SFKV 544

RESULT 2
OSGKT3
ID Q9GKT3 PRELIMINARY; PRT; 317 AA.
AC Q9GKT3;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE Hypothetical 34.7 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
TI TISSUE-CEREBELLUM CORTEX;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
KW EMBL; AB052146; BAB19001.1; -.
DR Hypothetical protein.
SQ SEQUENCE 317 AA; 34746 MW; 26C0DB80505BFCC4 CRC64;

Query Match 24.0%; Score 1532.5; DB 6; Length 317;
Best Local Similarity 93.4%; Pred. No. 2.1e-90;
Matches 297; Conservative 3; Mismatches 17; Indels 1; Gaps 1;

QY 898 MLOKSLSVIASAKEKTLGLAGKTQTAGVEERTKSKQPLPKDKETNRHNSDNTETKDP 957
Db 1 MLOKSLSVIASAKEKTLGLAGKTQTAGVEESAKSKQPLPKDKETNRHNSDNTETKDP 60
QY 958 PONSNSPAEPRKPQKSGIMQQRVNPTTANSDLNPGTQTKMKNFDIGVCPWEVYDLTPG 1017
Db 61 PONSNSPAEPRKPQKSGIMQQRVNPTTANSDLNPGATQMKDNFDIGVCPWEVYDLTPG 120
QY 1018 PVPSESKVKQHVSIIVASEMEKNPTFSLKEKSHHKPAAEVCQSNOKRDKAEVCLWESQ 1077
Db 121 PVPSESKVKQHVSIIVASEMEKNPTLUSLKEKSHHKPAAEVCQSNOKRDKAEVCLW 180
QY 1078 GQSILEDEKLLISKTPVLPERAKEENGOPRAANVCAGOSEELPPKAVASKTENENLNOI 1137
```

```
Db 181 GQSILEDEKHFISKTPVLQERAKEENGOPHAAKVCAGOSEELPPKAVASKTENENLNOI 240
QY 1138 GHQEKKTSSSEENVGSGYSSNNFQOPLTSRAEVCWPWEETPAQPNAGRSVALPASSALS 1197
Db 241 GHQEKKT-SFEENVGSGYSSNNFQOPLMSRAEVCWPWEETPAQPNAGRSVALPVSSALS 299
QY 1198 ANKIAGPRKEEIIDSFVK 1215
Db 300 ASKIAGPRKEEIVDITFKV 317

RESULT 3
OSVR40
ID Q9VR40 PRELIMINARY; PRT; 669 AA.
AC Q9VR40;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CG11923 protein.
GN CG11923
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadeau E., Center A., Chandra I.,
RA Cherry J.M., Crowley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003575; AAF50967.1; -.
DR FlyBase; FBgn0031642; CG11923.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1.
DR PROSITE; PS50259; G_PROTEIN_RECEP_F3_4; 1.
```



```
SQ SEQUENCE 669 AA; 75242 MW; FFE5721445DAB5F6 CRC64;
Query Match 9.0%; Score 571.5; DB 5; Length 669;
Best Local Similarity 26.1%; Pred.No. 2.5e-28;
Matches 195; Conservative 113; Mismatches 316; Indels 123; Gaps 27;
QY 110 GKWPALASAPSHLRALDTLTHATNFNLMYLNKSNREONLQDDLDLQALVMSLLEGEPP 169
||||| : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 6 GKWPC-RMFYDYTDIAEDAAQFIEFL-----SGKPNANTPIAID-----EP 47
QY 170 SISRAITSTSLSPAPQVFLQATRESRIILLQDLSSAPHLA-----NATLSEWTF 223
||| : : : : : | | : : : : : | | : : : : : | | : : : : : | | : : : : :
Db 48 --TRAESVRANGIAS-----YALNEDNLFAFAAPSIHTVYVVKFRDNVTIPPQV 98
QY 224 HG---LRRKWRPHLHRRGPNQGRGLGHSWRRKDGGLGKSHFKWSPPYLECEGNGSKPG 280
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 99 HNKAYLGSTWRE-----LCAAWNSTDGTQ-----EWGAPFDCNLLTRRWL 139
QY 281 WLVTLSAIYGLQPNLVPEFRGVMKYVDINLQKVIDQCSSDGW---FSGTHKCHLNSEC 337
| | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 140 WPPRISFS-----EHR-IKVVAFAIAADEVC-NDGLEEVFGRHGGCDRNTTC 187
338 MPKGL-GEVLGAYECICKAGFYHPCVLPVNNFRRRGPDQHSIGSTKDVSE--EAVVCLP 394
| | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 188 LUTENKPAATRDVYTCLCRESYLPN-STLQCFR-----GDRVELSEGYDNYSCIP 237
QY 395 CREGCPFCADDSPC--FVQE-----DKYLRLATISFQGLCMLLDVSMVYVHFRKAKSI 447
| | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 238 CPGGCTNCDNSGCLTFQEEVLNVNDACLRLLVAIVLGACILCCIVGLVIVFQRCKAI 297
QY 448 RASGLILLETILFGSLLLYFPVILYFEPSTFCILLRWALRGFATVYGTVKLHRYL 507
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 298 ASGMWTVLETILGIVLLYASVAVHFFPASTERCLLEPWLRELGFITCYGAILKLYRHL 357
QY 508 KYFLSRTAORIPWMTGGRVWRMLAVILLVVFVFLGWTSSVCONLEK-QISLIGQKTS 566
| | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 358 VDETRKAHRW-VLRDVLGLTGYMTFVAVICYMAAFTASSLDLLESAQLESREADT-- 414
QY 567 HLIFNMCLDRWDYMTAVAEFLFLMGVYLCYAVRVVPSAFHEPRYMAVAVHNELIISAI 626
| | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 415 ----NTPCLKWELVTQTEMLILFCGLHLSASRANTQFRERQFLVTLTLEFLVSS 470
QY 627 FHTIRVLSRQSDWMLMYFAHTHTVTVTIGLLLPK--FHSNPNPRDDIATEAYE 684
| : | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 471 FYFLREVLPENSPSAILLALFTRSLQTSNLFALGLFVFKLWYQKQGTSHD-----AG 524
QY 685 DELDMGRSGSYLNSSINSANSEHSLDPEDIRDELKLYAQLEYKKKMTNPNHLOKRR 744
| | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 525 QRLGGYAGLCGLGDPDIBELTISEMSPDIRAELKLYTOLEIMKNKTLRQDNPHISKRR 584
745 CSKKGLGRSIMRRITEIPETVSROCKEDKEGADHGTAKG-----TALIRKNPPESSGN 798
| : | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 585 GGRKAGHRRF-----SLOKKGSKALSAKHRSNKHQHDIEITEAPSTPDSVC 635
QY 799 TGKSKETLK--NRVPSLKKSHSTYDH 823
: : : | | : | | | | |
Db 636 SAEGPDTVAETISGVSHSLSHSWSH 662
RESULT 4
O45500 PRELIMINARY; PRT; 594 AA.
AC O45500;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE F39B2.8 protein.
GN F39B2.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RT Science 282:2012-2018(1998).
DR EMBL; Z92834; CAB07388.2; -.
SQ SEQUENCE 594 AA; 67920 MW; 80AF1FC165AB9070 CRC64;
Query Match 6.4%; Score 408.5; DB 5; Length 594;
Best Local Similarity 25.0%; Pred.No. 6.1e-18;
Matches 137; Conservative 89; Mismatches 192; Indels 131; Gaps 19;
QY 259 DKSHFKWSPPY-----LECENGSYKPGWLVTLSSAIYGLQPNLVPEFRGVM 304
||| | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 104 DKSSFOOIPPFDVSPFARRTHFYGLQNLCTGNHR--WLPRL--VFVSHDPN---NSNKA 156
QY 305 KYDINLQKVIDQCSSDGHGFWGTHKCHLNSECMPKGLGVLGAYECICKAGFYHPCVL 364
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 157 YLTQLDQDFDVMCS-----AVNCHTKCTWTVHGGL 187
QY 365 PVNNFRRRGPDQHSIGSTKDVSEAVVCLPCREGCPFCADDSPCFVQED-KYLRLAISF 423
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 188 RVS-----AKAC-----CNSDDVILCTDAERGRHMLLAF 217
QY 424 QGLCMLLDVSMVYVHFRKAKSIRASGLILLETILFGSLLLYFPVILYFEPSTFCIL 483
: | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 218 NSICALI-CIALIPVICYQKRQHEARGWALMELFLIGASILYSVLFDVFAPEYGCVC 276
QY 484 LRWARLLGATVYGTVKLHRVLKVFLSRTAQRIPMTGGRVWRMLAVILLVVFVFLG 543
| : | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 277 AVWLROIGFSIFGVSVLKIYRNLOEYRVKQHVSVREQDMKLYLAALMTITGLMA 335
QY 544 WTSVSVCONLEKQISLIGQKTSDLIFN---MCLIDRWDMYMTAVAEFLFLMGVYLCYA 599
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | |
Db 336 WT-----VGSWGDALWRTAWPQCLMQGHHVHWGVELLFLYAVRLCYK 380
QY 600 VRTVPSAFHEPRYMAVAVHNELIISAFHTIRVLSRQSDWMLMYFAHTHTVTVTI 659
| : | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 381 ARN--SDWLERWQFTVAVCLEAVITLNLRIYSIRNSGRADTLFIVSFVHLQTLTVSNI 438
QY 660 GULLIIPKFSHSSNNPRDDIATEAYEDELDMGRSG-----SYLNSSINSANSE-H 707
: | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 439 VIIVAPKFYLSNGEPPSRSSMTLG-----GHSGRAHPSLAKLRDNLINGTIDFAEVP 491
QY 708 SLDPEDIRDELKLYAQLEYKKKMTNPNHLOKRCCKGLGRS--IMRRITEIPETV 765
: | | | | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : | |
Db 492 DMNPEDIRAEKRVYTLRMVYKLNLYQDNPHISKARGKKWSDKNTKATRRIS-IP--- 547
QY 766 SRQCKEDK 774
| | : | |
Db 548 --SCSPQTK 554
RESULT 5
Q9VKA3 PRELIMINARY; PRT; 1677 AA.
AC Q9VKA3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CG17215 protein.
GN CG17215.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Balwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Delcher A., Deng Z., Guan P., Harris M.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003635; AAF53175.1; -.
DR FlyBase; FBgn0032418; CG17215.
DR InterPro; IPR002524; Cation_efflux.
DR InterPro; IPR000337; GPCR_Mgr.
DR InterPro; IPR000834; Zn_carbOxep.
DR Pfam; PF01545; Cation_efflux; 1.
DR TIGRfams; TIGR01297; CDF; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; UNKNOWN_1.
DR PROSITE; PS50259; G_PROTEIN_RECEP_F3_4; 1.
SQ SEQUENCE 1677 AA; 188028 MW; F50A8D282A8E3B6E CRC64;

Query Match
Best Local Similarity 4.8%; Score 307; DB 5; Length 1677;
Matches 124; Conservative 98; Mismatches 238; Indels 118; Gaps 19;

QY 443 KAKSTRASGLLELITLFGSLLLYPPVILYPEPSTFCRILLRWLRLLGFAVYGVTVLK 502
DB 1157 KAGRRGSDVAI-----FPYL-----DTWCIAKWRHMGFCITYSLMLK 1199

QY 503 LHRVLKVLFSRTAQRIPYMGVRMRMLAVLLVVFVFLIGHTSSVCQNLKQISLIGQG 562
DB 1200 TWRVSLTVRKVSAHKIK-LNDQOQLQWMPILLVMLIYLGWTISATPTAEVILD----- 1253

QY 563 KTSDDLIFNMCLIDRDYMTAVAEFLFLWGYLYCYAVPTVPSAEHPYMAVAVHNELI 622
DB 1254 --OSQKFKQCQSYNMWDHSLAIGEYFFFLAWGRVYCNVNAESLYNEARLISYATYNIAL 1311

QY 623 IS---AIFTRIFVLASRLQSDMLMLFAHTHTLTVTTVIGLLLPKFKSHSSNNPRDDTA 679
DB 1312 VNIAWVHVHMLFPNAG---PPKYMLGFVTRQLSTTTTIALVFGPKILR-----V 1359

QY 680 TEAYEELDMGRSGSYLNSINSNSAWSEHSL-----DPEDIRDELKLYAQ 724

```

```

Db 1360 FRQGGDKWQ-----KAKVRISITASFSLNGVLVPEESPOLYQNEELKEQVQKLAHQ 1412
QY 725 LEIYRKKKMITNPHLOKKRCKSKKGLGRSIMRRITPEITVSPQCSKDEKGDHGTAKG 784
Db 1413 IFBMKTVHQMNNRHUK-----PKPGGYFTITSTFOAPYSKNTVSTAQTQTKDENSVDK 1468
QY 785 TALIRKNPPSSGNTGSK--BETLKNRVFSLKKSHSTYDHRDQTEESSSLPTESQEEE 842
Db 1469 CSI-----EIDGQGRITLVEDAIE--FNF-----HLAD--TEKGLIVQEDRPG 1509
QY 843 TTENSTLESLSGKKLTQKLKDESAESTESVPLVCKSASAHNLSSEKKTGHPRMTLSQKS 902
Db 1510 SDEEEAALIAQFRFLFAPILDDSD-----LNLVYQLNDLDDTEHVRHQTVAQ 1556
QY 903 LSVIASAKERTGLGACKTOTAGVEERTKSKQPLPKDKETNRHNSNDTETKDPAPQNSN 962
Db 1557 MNDLTSSEETM-----VTQVNSPPPPGVGVELLLFISSDSST-----ASSLY 1600
QY 963 PAEPRKPKQSGI--MKQQRVNPPTTANSDLNPGTQOMK 998
Db 1601 AHTTSPAHPSGVLLMPQNLSPLESPLISGSDAITEQVE 1638

RESULT 6
Q9WLP3 PRELIMINARY; PRT; 264 AA.
AC Q9WLP3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG18678 protein.
GN CG18678.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Balwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Delcher A., Deng Z., Guan P., Harris M.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

```



Db 509 LSSVAAFPVICTARAWILMAGFSLSGAMFSTWRVHSIFTDLKLNK-KVVDYQLFMV 567  
Qy 531 AVILLVWFLLGW-TSSVCQNLKQISLIGOGKTSDDLIFNMCLIDRWYMTVAEFLF 589  
Db 568 GVLLAIDIAITITWQADPFYRETQLEPLHHENIDDLVPIPENECQSEHMTIFVSIY 627  
Qy 590 -----LLMGVLYCAVTRVP-SAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDW 642  
Db 628 AYKGLLVFGAFWETRHVSIPALNDSKHIGFSYVNVFITCLAGAAISLVLSDRKDLVF 587  
Qy 643 MMLYFAHTLTVITIGLLIPKFSHSSNNPRDDIATEAYEDELDMGRSGSYLNSINS 702  
Db 688 VLLSFP--IIFCTTATLCLVFPVKLVELKRNPGQVVDKRVATLRPMKNGR----- 737  
Qy 703 ANSEHSLDPEDIRDELKLYAQLEIYKRRKMTN-NPHLQKRCCKGLGRSMTMRITEL 761  
Db 738 -----RDS-----SYCELEQRDLRVKNTNCRFKAHKMEKENELQALIRKL 778  
Qy 762 PETVSQR-----CSKEDKEGADHGTAKGTALIRKNPPESSNGTGSKEETLKNRVFSLK 815  
Db 779 PE--ARKWIDGVTCTGSGNVGSELEPILNDDIVRLSAPPVRRMPSTTEVTMTSVDV 836  
Qy 816 KHSYTDH--VRDQTESSSLPTESQEEETEN-----STLESGLKGLTKQKLEDSE 866  
Db 837 STHVMDNSFVSQSTVMAPSLPPKKKQSIIVEHHSHAPAPTMWOPIQOOLQOHLQHQ 896  
Qy 867 AESTESVPLVCKSASAHNLSSEKKTCHPRTSMLOKLSVIAKAKETLGLAGKTOTAG-- 924  
Db 897 MOOQH-----LOOQOHHQOQOQOHHHRLKRNVSQAOTD---NIGSITTAGKR 948  
Qy 925 -----VEERTK-----SOKPLPKDETNRNHSNSDNTETKDPAPONSNAEP 967  
Db 949 SGGDCSSMERERQSTASRYHDSGQTPTAPKYSSSHRNSNTNISTQSSELSNMCPHXP 1008  
Qy 968 RPKQSGIMKQORVNPPTANSDLNPGCTQMKDNFDIGEVCPWEVYDL-----TPGPVP 1020  
Db 1009 STP--AVIK-----TPTASDHRTTSMGSAKSNFVVSQSDLDWDTLTHAKORQSPR 1061  
Qy 1021 SESKVKHVSIVASEMEKNPT-----FSLKEKSHHKPKAAEVCQSNOKRIDKAEVCLW 1074  
Db 1062 SPQCAEHGGHGMYPDNTTSPQRSVSEKRNKRHPK-----OKGTVC-- 1107  
Qy 1075 ESQOSILEDEKLLISKTPVLPRAKEENGQPPRAANVCAGQSEELPPKRAVASKTENENL 1134  
Db 1108 QSETDSERERDPPNSQPCVQPRKYS-----RSSNIQHAHHSSPNVADK---QRS 1157  
1135 NQIGHQEKK---TSSEENVRGSYNSNNFQOPLTSRA-----EVC 1173  
Db 1158 RQKGQDSSIIYGASSETELLEGETALIPFRKLLTEKSPNRCRSVAVGOSCP 1209

RESULT 8  
Q9BML6

ID Q9BML6 PRELIMINARY; PRT; 1220 AA.  
AC Q9BML6;  
DT 01-JUN-2001 (TremBLrel. 17, Created)  
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Metabotropic GABA-B receptor subtype 2.  
GN GABA-B-R2 OR BCDNA:GH07312 OR CG6706.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21097320; PubMed=11168554;  
RA Mezler M., Muller T., Raming K.;  
RT "Cloning and functional expression of GABA-B receptors from  
RT Drosophila.";  
RL Eur. J. Neurosci. 13:477-486(2001).

DR EMBL; AF318273; AAK13421.1; -.  
DR FlyBase; FBgn0027575; GABA-B-R2.  
DR InterPro; IPR001828; ANF\_receptor.  
DR InterPro; IPR000337; GPCR\_Mgr.  
DR Pfam; PF00003; 7tm\_3; 1.  
DR Pfam; PF01094; ANF\_receptor; 2.  
DR PROSITE; PS0259; G\_PROTEIN\_RECEP\_F3\_4; 1.  
KW Receptor.  
SQ SEQUENCE 1220 AA; 137976 MW; 2B33DA2CA1BDA8B CRC64;

Query Match 3.8%; Score 243.5; DB 5; Length 1220;  
Best Local Similarity 17.5%; Pred. No. 6.4e-07;  
Matches 212; Conservative 188; Mismatches 321; Gaps 47;

Qy 72 OKLAEEVPMVASYLYTGDHQKLRANCSG-RYELAGLPKWPALASAHPSLHRLDTIT 130  
Db 216 KKLREK---DVRILIGNFNEHFAKFAFCAYKLDYGRAYQWLMA-----T 259  
Qy 131 HATNFMVLMQNSKREQNLQDLDWYQALVMSLGEPSISRAAITFTSDLSAPAPQV 190  
Db 260 YSTDWNVYTDSECVEE-----IATALEG-----ALLVDLLPLSTSGDIT 300  
Qy 191 FLOATREESRILLQDLSAPHLANATLETFWHLGRK--WRPHLHRRGPNQGRGLGH 248  
Db 301 VAGITADEYLVEYDLRGT-----EYSRFHGYTYDGIWAAALAIQYVAE----- 344  
Qy 249 SWRRKDLGGDKSHF-----KWSPPYLEC-ENGSKYKPGWLVTLSAIYGLQPNLYPEFRG 302  
Db 345 --KRELL---THFDYRVKDWESVLEALRNTSEG---VTGPVRFYNN----- 386  
Qy 303 VMKVDINLQKVDIDQCSSDGMFSGTHKLNNSSECMKGLGLFVLGAYECICKAGFYHFG 362  
Db 387 -RKANLLNQFOLGQWKEIGEVH-SQKSHLDLSLCKPVKVG----- 426  
Qy 363 VLPVNNFRRRQDQIHSGSTKDVSEAYVCLPCRGCPFCADDSPCFQVEDK-----YLR 417  
Db 427 -----KTPPK-----DRTLVIHESQVNPITIV 450  
Qy 418 LAIISFOGLCMLLDFVSMVYHFRKAKSIRASGLLILLETILFGSILLYFPVVLVYFPS 477  
Db 451 SASASVIGVIATVFLAFNIKY--RNQRIKMSPHLNNLIIVGCMITVLSIIFGLDPT 508  
Qy 478 -----TFRCLLWRLGLFATVYGTVTLKHLRVKLVFLSRTAQRIPYMTGGVRML 530  
Db 509 LSSVAAFPVICTARAWILMAGFSLSGAMFSTWRVHSIFTDLKLNK-KVVDYQLFMV 567  
Qy 531 AVILLVWFLLGW-TSSVCQNLKQISLIGOGKTSDDLIFNMCLIDRWYMTVAEFLF 589  
Db 568 GVLLAIDIAITITWQADPFYRETQLEPLHHENIDDLVPIPENECQSEHMTIFVSIY 627  
Qy 590 -----LLMGVLYCAVTRVP-SAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDW 642  
Db 628 AYKGLLVFGAFWETRHVSIPALNDSKHIGFSYVNVFITCLAGAAISLVLSDRKDLVF 687  
Qy 643 MMLYFAHTLTVITIGLLIPKFSHSSNNPRDDIATEAYEDELDMGRSGSYLNSINS 702  
Db 688 VLLSFP--IIFCTTATLCLVFPVKLVELKRNPGQVVDKRVATLRPMKNGR----- 737  
Qy 703 ANSEHSLDPEDIRDELKLYAQLEIYKRRKMTN-NPHLQKRCCKGLGRSMTMRITEL 762  
Db 738 -----RDS-----SYCELEQRDLRVKNTNCRFKAHKMEKENEL- 770  
Qy 763 ETVSRQCKEDKEGADHGTAKGTALIRKNPPESSNGTGSKEETLKNRVFSLK----- 816  
Db 771 QALIRKLGPARKWIDGVTCGT-----GSNVGSELEPILNDDIVRLSAPPVRR 819  
Qy 817 -----SHSYDHR-----DQTESSSLPTESQEEETEN-----STLE 850  
Db 820 MPSTTVTMTSVDVSTSTHVMDNSFVSQSTVMAPSLPPKKKQSIIVEHHSHAPAPTM 879  
Qy 851 SLSGKKLTKQKLEDSEABESTSVPLVCKSASAHNLSSEKKTCHPRTSMLOKLSVIAK 910  
Db 880 QP1OQOQLQOHLQHQOQOQH-----LOOQOHHQOQOQOHHHRLKRNVSQAOTD 934

[illegible]

```
Db 378 QENPKYKNTCSQWTLRTOHQVDSKMGFVINAISYMAVGLHNQSLQCPGYVGL----- 431
Qy 310 LQKVIDQCSGDFSGTHKCHLNSECPKIGLGVLCAYECICKAGFY----- 359
Db 432 -----CDA-----MKPIDGRKLL-----ESLMTNFTGSGDMILFD 463
Qy 360 ----HPGVLPVNNFRRRGPD--QHI-----SGSTKDVSEE----- 388
Db 464 ENGDSGRYEINFKMGKDYEDYINGSWDNGELKMDDEIWSKNNIIRSVCEPCK 523
Qy 389 -----AYVCLPCREG-----CPFCA-----DSPCFVQEDKYLR--- 417
Db 524 GOIKVIRKEVSCCWTCTPCKENYVFDEYTCACQOLGSWPNDELTCGLIPQVYLRWGD 583
Qy 418 ---LAIISFQGLCMLLDFVSMVYVHFRKAKSIRASGLILLTILFSGLLLPVVVILYF 474
Db 584 PEPIAAVFACLGLLATLVTAIFTMYRDTVPVVKSSRELVCYIILAGICGLVCTFCLIA 643
Qy 475 EPSTFCILLRWARLIGFATVYGTVTLKLRVLYK-----FLSRTAQRIPIY 520
Db 644 KQOIYCYLQIRIGISPAWSALVTNRTARILAGSKKICKKPRFMSACAQ---- 599
Qy 521 MTGGRVMRLAVILLVVFVFLIGWTSVQCNLEKQISL-----IGQKTSDHILFNMC 573
Db 700 -----LVIAFILI-----CIQLGIIVAFIMEPPDIMHDYPSIREVYLIC 739
Qy 574 LIDRWDMYATA-BEFLWGLWYLYAVYRTVPSAHEPRYMAVAVHNELIISAIHTIRF 632
Db 740 NNTNLGVVTPGLYGLLLISCTFYAFKTRNVPANFNAKYIAFTYTTCIIMLAEVPIYF 799
Qy 633 VLASRLQSDWMLMLYFAHTHLVTVTIGLLLPK-----FSHSSNNPRDDIATE-AYEDEL 687
Db 800 -----GSNYKIITWCFVSLSATVALGCMFVPKVYIILAKPERNVRSAFTSTVVRHV 853
Qy 688 DMGRSGSYL--NSSINSAMSEHSLDPEDIRDELKLYAQLEIYKRRKMTNNPHLQKRC 745
Db 854 GDGKSSAASRSSLVNLWKRSGSGETLR-----YKGRRLA--PHKSEIEC 898
Qy 746 -SKKGL--GRSIN--RRITPIETVSPQCKEKGADHGTAKGTALIRKNPPESSGN 798
Db 899 FTPKSGMNGGRATTSSESVCIPECNRSSESRDEKE-----VPVKEDAL 943
Qy 799 TGKSEETLKNRVFSLKKSHSTYDHRDQTESSLSPTESQBEETENSTLESKGLIT 858
Db 944 TCK-----KTCNCVSLIVPODCOLODLLKHSNCKSVSWAQNKSXRGALHW 990
Qy 859 QKL-----KEDSEASTESVPLVCKSAAHNLSSBKKTGHPRTSMQLKSLSVIASAKET 913
Db 991 ORLSITHINKENPNQTAVIKPSKSTDSRSHSS---ATFPETS--AKTLYDVSEAEQ- 1044
Qy 914 LGLAGKTQTAGVEERTKSKPLPKDKEFNHNHNSDNTETKDPAPQ-NSNPAAEPKPKQ 972
Db 1045 -----YPAQYRPQTPSPISVSHRTASVS-----RTEDDAPTQSEPPQRSSSQG 1090
Qy 973 SGIMKO--ORVNPTTAN--SDLN-----PGT--TOMKDNFDIGEVCPWVYDLTPGP 1018
Db 1091 S-LMQISVSVTRTANISELSNMLSTATPGTWATPLCSSY-----LIPREIQ--LPTT 1143
Qy 1019 VPSESKVQKHVS----VASEMEKNPTFSLKESKSHHKPAEAVCOQSNOKRIDKAEVCLW 1074
Db 1144 MTTFAEIQPLPSIEVNGASQSAKQSGSVKEGTAEFTSAK----- 1184
Qy 1075 ESQGSILEDEKLLISKTPVLPRAKENGQOPRAANV-----CAGQSEELPPKAVASKT 1129
Db 1185 -----QDLEELVALTPPSPFRSDIDSASPSPVSESALCIPSPPKYDILLIRDT 1236
Qy 1130 ENEN 1133
Db 1237 QSSS 1240
RESULT 11
```

```
Q9BML5
ID Q9BML5 PRELIMINARY; PRT; 1305 AA.
AC Q9BML5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Putative metabotropic GABA-B receptor subtype 3.
GN GABA-B-R3 OR CG3022.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21097320; PubMed=11168554;
RA Mezler M., Muller T., Raming K.;
RT "Cloning and functional expression of GABA-B receptors from
RT Drosophila.";
RT Eur. J. Neurosci. 13:477-486(2001).
DR EMBL; AF318274; AAK13422.1; -.
DR Flybase; FBgn0031275; GABA-B-R3.
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR InterPro; IPR000005; HTHaraC.
DR Pfam; PF00003; 7tm.3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS0259; G_PROTEIN_RECEP_F3.4; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
SQ RECEPTOR.
KW RECEPTOR.
SQ SEQUENCE 1305 AA; 143701 MW; 8BFA80F0E9BEADD CRC64;
```

Query Match 3.6%; Score 232; DB 5; Length 1305;  
Best Local Similarity 18.6%; Pred. No. 3.9e-06;  
Matches 228; Conservative 183; Mismatches 396; Indels 416; Gaps 59;

```
Qy 63 RSTGCTILAQKLAEBVPMVDVASYLYTGDSHQLKRNCSGR-----YELAGLPGKWP--A 114
Db 344 RETDTRIIGSFQSELAPQILCEAY-----RUMFGADYAWILHESMGAPW-WPQOR 394
Qy 115 LASAHPSLRALDTL-----TH-----ATNFLNVLQSNKSRQNLQDDLDWYQALVM 162
Db 395 TACSNEHLQAVENLIVVSTHNSIVGNVSVYSGLNHMFNSOLRKQSAQ----- 443
Qy 163 SLLEGEPSIRAAITFTSDLSAPAPQVLOATRESR-----ILLQDLSS 208
Db 444 --FHQDQGF-----SGYGPRIASIAATQSDSRRRRRRVYVGTSGGHLFPEAISQ 490
Qy 209 SAPHLANATLETENPHGLRRKWRPHLHRRGPNQGRGLGHSWRRKDGKSHKFWSP 268
Db 491 YAPQ-----TYDAVMAIALALRAAEHWRNEQS-----KUDGFDYTRSDMAWE-- 535
Qy 269 YLE-----CENGSKYKPGWLVTLSAIYQLQPNLVPEFRGMKVDINLQKVD 314
Db 536 FLOQMGKHLFGLVSGPVSFSGPDRVG-----TTAFYQIQ-----RCL-----LEPVA 577
Qy 315 IDQCSSDQWFGSTGHKCHLNSECPKIGLGVLCAYECICKAGFYHPGVLPVNN--FRRR 372
Db 578 L-----YYPATDALDPRCPRPVK-----WHSQGVPIAKRVFKLR 613
Qy 373 GPDQHSISTKDVSEAEVCLPCREGCPFCADDSPCFVQEDKYLRLAISFQGLCMLLDF 432
Db 614 -----VATIAPL-----AFYTIATLSSVGIATAITF 639
Qy 433 VSMLVVYHFRKAKSIRASGLILLETILFSGLLLPVVVILYFEPST-----FRCELLR 485
Db 640 LAENL--HERKLAIKLSSPKLSNITAVGCFVYATVILLGLDHSHTLPSAEDSFATVCTA 697
Qy 486 WARLL--GFATVYGTVTLKLRVLYKVFLSRTAQ-----RIPYMTGGRVMRLAVI 533
Db 698 RYVLLSAGFSLAFGSMFAKTYRVHRIF--TRTGSVFKDKMLQDIQLILLVGG---LLLVD 753
```





```
Db 689 DSFATVCTARVYLLSAGSFAGSMFAKTYRVHRIF-TRTGSVFQKMLQDQLILLVGG 747
Qy 525 RYMRMLAVILLVWFLLGTWTSVCQNLKQLSLGQKTSDFLIF-NCML---IDR 577
Db 748 ---LLLVALLVTLKVVDPMERHHLNLTLEIS-----ATDRSVYQPVQVEVRSQHTQT 799
Qy 578 WDMYTAFAEFLLVGVLYCYAVRVP-SAPHEPRYMAVAVHNELIISAIFHTIREVLAS 636
Db 800 WLSVLYAYKGLLVGVYMAWETRVKIPALNDSYIGVSVVVTSAI-----VVVLA 854
Qy 637 RIQSDWMLMLYFAHPL---TVVTYIGLLLPKFHSSNPRDDIATEAYEDELDMGRSG 693
Db 855 NLISERVTLAFITITALILTSTTATLCLLFIKP-----887
Qy 694 SYLNSINSANSEHS-LDP-----EDIRDELKLYAQLIYKRRKMI 734
888 -----LHDIIWARNDIIDPVIHSMGLKMECNTRRFVDDRRRELOYRVQNRVYKKE--- 938
735 TNNPHLOKKRCSKGLGRSIMRRITEIPETVSROCSEKDEGADH-----GTAKGT 785
939 -----IQALDAIRKLERLLSGLTFTTTSSSTSL--LTGGGHLKPELVVTSISQTP 991
Qy 786 ALIRKNPPSSGNTGKSKEETLKNRVFSL-----KSHSTYDHRDQTESSSLPTESQ 839
Db 992 AASKNRTPSIG-----ILPNLLSVLPVPRASWPSAEYMQIPMRRSVTFASQPQ 1043
Qy 840 BEETTENTLESLSGKLTQKLKEDSEAEVSTESVPLVCKAS-----AHNLSEKKTG 892
Db 1044 LEE-----ACLPADQLINLRHLAQHQATEKG 1070
Qy 893 -----HPRTSMLOKLSVIASAKETGLAGKTQTAGVEERTKSKQPLPKDKETNRN 944
Db 1071 LINLRGIFRTSSNKG-STASLADQGLKAAPKASH-WGLETRL-----TPSSQTASCN 1123
Qy 945 HNSDNTETKDPAPONSNAPEERPKQSGIMKQORVNPTTANSDLNPGTGMKD-NF-D 1002
Db 1124 AIYNNPNQDISPSEASSHPNGHLKPIHRGSL-----TKSGTHLDHLT---KDPNPLP 1173
Qy 1003 IGEVCPWEVDLTGCPVPSESKVKHVIASEM-----EKNPTF-----SLREKSHHK 1051
Db 1174 IPTISGGGQDQTLG-----GKYVLLKETVNFQLPNSNRPRSVQOQPSLREVRGS 1225
Qy 1052 PK-----AAEVCQSQNOKRIDKAEVCLWESQCSILEDEKLLISKTPVLPRAKEE--NG 1104
Db 1226 PRFPHRILPPCTSL-----ALAESEDRP--GDSILGCKSIPRLISQOATSG 1273
1105 GQPR-----ANVCAGOSEELPPKAVASKTE 1130
Db 1274 GTWKSMTAGKSRSLSGDSQEEQAPANGTE 1305
```

## RESULT 13

```
Q8SY55
ID Q8SY55 PRELIMINARY; PRT; 1514 AA.
AC Q8SY55;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE GH09335p.
GN CG6004.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
```

```
RA Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY075323; AAL68190.1; -.
SQ SEQUENCE 1514 AA; 157140 MW; 1FFC4B0664105AD5 CRC64;

Query Match 3.5%; Score 225; DB 5; Length 1514;
Best Local Similarity 20.5%; Pred. No. 1.3e-05;
Matches 129; Conservative 91; Mismatches 281; Indels 128; Gaps 22;

Qy 633 VLARLQSDWMLMLYFAHPLTVTVTIG-----LL-----LIPKFHSSNPN----- 674
Db 235 IISSTEGSW-----ESHISTDSSIGSKVESLLEALYSLIQESSSSSPVSNPEST 287
Qy 675 ---RDDIATEAYEDELDMGRSGS-----YLSNSINAWSEHSLDPEDIRDELKLYAQ 724
Db 288 GATDDSSSTESLPDSTQESSSESPEVSFELSTEATNESSSESLPNSSTQDSSSTETS 347
Qy 725 LEIYKRKKMIITNPHLOKKRCSKGLGRSIMRRITEIPETVSROCSEKDEGADHGTAKG 784
Db 348 FQTESTTDATDESSSTESQPDSTTQESSS-----STEGPLSTESSTAVTDQSSSTESQDS 403
Qy 785 TALIRKNPPSSGNTGKSKEETLKNRVFSLKSHSTYDHRDQTESSS-----LPTESQ 839
Db 404 TQESSSSTEGPLSTESSTEA-----NESSSTESQDSTTQESSSSTEGPLSTESS 455
Qy 840 BEETTENTLESLSGKLTQKLKEDSEAE-STESVPLVCKASAHNLSEKKTGHPTSM 898
Db 456 TEATNESSSTES-SQDSTTQESSSTEGPLSTES-----STEATNESSSTESSQDSTT- 507
Qy 899 LOKLSVLIASAKETGLAGKTQTAGVEERTKS-----OKPLPKDKETNRN 945
Db 508 -QES-----SSSEGPLSTESSTEAATNESSSTESSQDSTTQESSSSTESPLSTESSTEA 562
Qy 946 SNSDNTETKDPAPONSNAPEERPKQSGIMKQORVNPTTANSDLNPGTGMKDNFIDG 1005
Db 563 SSSTESSQDSTTQESSSTEDPLSTESS-----TEATNESSSTESSQDSTTQ----- 609
Qy 1006 VCPWEVDLTGCPVPSESKVKHVIASEMEKNPTTSLREKSHHKPKAAEVCQSQNOKR 1065
Db 610 ---ESSSTEGPLSTESSTEGSNESSSTESSQDSTTQKSSSTESPLSTEPSTEANESS 665
Qy 1066 IDKAEVCLWESQCSILEDEKLLISKTPVLPRAKEE-----GGOPRAANVCAGOSEE 1119
Db 666 STES-----SQDSTTQESSSSTEGPLSTEPSTEANESSSTESSQDSTTQESSSSSE- 716
Qy 1120 LPPKAVASKTENENLNGHOKKTSSEENVRSYNSNNFQOPLTSRAEVCWEFETP 1179
Db 717 -GPLSTESSTE-----ANESSSTESSQDST--TQESSSSTESPLST-----EPSTE 759
Qy 1180 AQPNAGRSVALPASSALSANKIAPRKEE 1208
Db 760 ANESSSTESSQDSTTQESSSSTEGPLSTE 788
```

## RESULT 14

```
Q76891
ID Q76891 PRELIMINARY; PRT; 5327 AA.
AC Q76891;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE EG.49P4.1 protein.
GN FUTSCH OR EG.49P4.1 OR CG3064.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RT "Sequencing the distal X chromosome of Drosophila melanogaster.";
```





Db 1064 -----PCNGLSLYPPPPQ-----HLQMLPLQLS-----TFREPISEP--G 1100  
QY 1056 EVCQSQNQKRIDKAEVCLWESOG---QSILEDEKLLISKTPVLPERAKEENGQPPAANV 1112  
Db 1101 EDDDDSSERFKLLQEFVYERECNTEEDLEBEEDLPAASKLTPEDSPALTPPSPPRDSV 1160  
QY 1113 CAGQSEELPP--KAVASKTENENLNQIGHQEKKTSS 1147  
Db 1161 ASGSSVPSSPVSESVLCTPPNVTYASVILRDYKQSS 1197

Search completed: February 15, 2003, 08:30:53  
Job time : 86.6488 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2003, 00:11:08 ; Search time 23.4427 seconds  
(without alignments)  
1524.948 Million cell updates/sec

Title: US-09-775-181-2  
Perfect score: 6382  
Sequence: 1 MGAMAYPLLLCILLIAQLGLG.....LSANKIAPKKEIWDSEKVF 1215

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	220.5	3.5	1199	1 US-08-041-538-2	Sequence 2, Appli
2	220.5	3.5	1199	1 US-08-463-642-2	Sequence 2, Appli
3	220.5	3.5	1199	1 US-08-455-602-2	Sequence 2, Appli
4	220.5	3.5	1199	2 US-08-465-157-2	Sequence 2, Appli
5	220.5	3.5	1199	5 PCT-US91-09422-2	Sequence 2, Appli
6	220	3.4	1219	2 US-08-687-289A-6	Sequence 6, Appli
7	204	3.2	1180	4 US-08-660-148-2	Sequence 2, Appli
8	200	3.1	1180	1 US-08-486-270-8	Sequence 8, Appli
9	200	3.1	1180	3 US-09-153-757-8	Sequence 8, Appli
10	200	3.1	1180	4 US-08-367-264-8	Sequence 8, Appli
11	197.5	3.1	1212	4 US-08-660-148-5	Sequence 5, Appli
12	193.5	3.0	1212	1 US-08-486-270-10	Sequence 10, Appli
13	193.5	3.0	1212	3 US-08-367-264-10	Sequence 10, Appli
14	193.5	3.0	1212	4 US-09-153-757-10	Sequence 10, Appli
15	193	3.0	1180	1 US-08-072-574-8	Sequence 8, Appli
16	193	3.0	1194	4 US-08-538-526-1	Sequence 1, Appli
17	190	3.0	1056	2 US-08-687-289A-8	Sequence 8, Appli
18	186.5	2.9	1212	1 US-08-072-574-10	Sequence 10, Appli
19	186	2.9	1056	2 US-08-687-289A-7	Sequence 7, Appli
20	178	2.8	571	4 US-08-961-083-4	Sequence 4, Appli
21	177	2.8	906	1 US-08-486-270-2	Sequence 2, Appli
22	177	2.8	906	3 US-08-367-264-2	Sequence 2, Appli
23	177	2.8	906	4 US-09-153-757-2	Sequence 2, Appli
24	174	2.7	906	5 PCT-US91-09422-17	Sequence 17, Appli
25	173.5	2.7	1964	2 US-08-790-912-3	Sequence 3, Appli
26	173.5	2.7	2052	2 US-08-790-912-2	Sequence 2, Appli
27	172.5	2.7	1079	1 US-08-485-588-8	Sequence 8, Appli

Query Match 3.5%; Score 220.5; DB 1; Length 1199;  
Best Local Similarity 18.6%; Pred. No. 3e-09;

28	172.5	2.7	1079	1 US-08-484-565-8	Sequence 8, Appli
29	172.5	2.7	1079	2 US-08-480-751-8	Sequence 8, Appli
30	172.5	2.7	1079	3 US-08-943-986-8	Sequence 8, Appli
31	172.5	2.7	1079	3 US-08-353-784-8	Sequence 8, Appli
32	172.5	2.7	1079	3 US-08-484-719B-8	Sequence 8, Appli
33	172.5	2.7	1079	4 US-08-484-159-8	Sequence 8, Appli
34	170	2.7	877	1 US-08-486-270-12	Sequence 12, Appli
35	170	2.7	877	3 US-08-367-264-12	Sequence 12, Appli
36	170	2.7	877	4 US-09-153-757-12	Sequence 12, Appli
37	168.5	2.6	1848	4 US-08-296-791-6	Sequence 6, Appli
38	168.5	2.6	1848	5 PCT-US95-10661A-6	Sequence 2, Appli
39	168	2.6	1863	2 US-08-603-753D-2	Sequence 2, Appli
40	168	2.6	1863	4 US-09-099-753-2	Sequence 2, Appli
41	168	2.6	1863	4 US-08-986-106-2	Sequence 2, Appli
42	168	2.6	1863	4 US-09-007-678B-49	Sequence 49, Appli
43	165	2.6	877	1 US-08-072-574-12	Sequence 12, Appli
44	163.5	2.6	879	4 US-08-794-158-2	Sequence 2, Appli
45	163.5	2.6	905	1 US-08-072-574-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-041-538-2  
; Sequence 2, Application US/08041538  
; Patent No. 5385831  
; GENERAL INFORMATION:  
; APPLICANT: Mulvihill, Eileen R  
; APPLICANT: Hagen, Frederick S  
; APPLICANT: Houamed, Khaled M  
; APPLICANT: Almers, Wolthard  
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE  
; NUMBER OF SEQUENCES: 15  
; NUMBER OF INVENTION: RECEPTORS  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend  
; STREET: Steuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/041,538  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/648,481  
; FILING DATE:  
; APPLICATION NUMBER: US 07/626,806  
; FILING DATE: 12-DEC-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-6-1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-467-9600  
; TELEFAX: 206-623-6793  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1199 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-041-538-2

```

US-08-463-642-2
; Sequence 2, Application US/08453642
; Patent No. 5721107
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R
; APPLICANT: Hagen, Frederick S
; APPLICANT: Houamed, Khaled M
; APPLICANT: Almers, Wolfgang
; TITLE OF INVENTION: G PROTEIN C

```

Db 875 -----IFRRKKPGAGNANSNGKSVSWSEPPGGRQAPKQGHVWQRLSVHVKT 919  
Qy 765 VSRQSKEDKADGHTAGTALIRKNPSSGNGTKSKEETLKRNVSFLKSKSHSTYDHY 824  
Db 920 NETACNQ-----TAVIK---PLTKSYQSGSKSLTFSD----- 948  
Qy 825 RDQTESSSLPTESOEETSTLESLSGKLTQKLEDSEASTESVPLVCKKSASAHN 884  
Db 949 -----ASTKTLNVEEDNTPSAHFSPSPSPSMVVR 980  
Qy 885 LSSEKKTGHPRMTSLQKSLSVIASAKEKTLGLAGKTQAGVEERTKSKQKPLPKDKETNRN 944  
Db 981 RGPVATTPLPPLHL-----TAETPLFLADSVIPKGL-----PPPLPQOQ----- 1021  
Qy 945 HNSDNTETKDPAPQNSNPAEPRKQSGIMKQORVNPPTANSDL-----NPGTTQ 996  
Db 1022 -----POQPPQPPQPK--SLMDQLQGVVTFNGSGIPDFHAYLAGPGTP- 1065  
Qy 997 MKDNFDIGEVCPWEYDITPGVPSE--SKVQKHYSIVASEMEKNPTFSLKESHHKPKA 1054  
Db 1066 -----GN-----SLRSLYPPPPPHQLMPLHLSTFQEEISIPGEDIDDDSS----- 1108  
Qy 1055 AEVCOOSNOKRIDKAEVCLWESQ---OSILEDEKLLISKTPVLPRAKEENGOPRAAN 1111  
Db 1109 -----ERFKLQEFYIEREGNTEDELEDEEDLPTASKLTPEDSPALTPPSPFRDS 1159  
Qy 1112 VCAGOSEELPP--KAVASKTENENLNQIGHQEKTSSS 1147  
Db 1160 VASGSSVPSPVSESVLCTPPNVTYASVILRDYKQSSS 1197

RESULT 3

US-08-455-602-2  
; Sequence 2, Application US/08455602  
; Patent No. 5747267  
; GENERAL INFORMATION:  
; APPLICANT: Mulvihill, Eileen R  
; APPLICANT: Hagen, Frederick S  
; APPLICANT: Houamed, Khaled M  
; APPLICANT: Almared, Wolfhard  
; TITLE OF INVENTION: G PROTEIN COUPLED GLUTAMATE  
; TITLE OF INVENTION: RECEPTORS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend  
; STREET: Stuart Street Tower, One Market Plaza  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94105  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/455,602  
; FILING DATE: 31-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/672,007  
; FILING DATE: 18-MAR-1991  
; APPLICATION NUMBER: US 07/648,481  
; FILING DATE: 30-JAN-1991  
; APPLICATION NUMBER: 07/626,806  
; FILING DATE: 12-DEC-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parmelee, Steven W  
; REGISTRATION NUMBER: 31,990  
; REFERENCE/DOCKET NUMBER: 13952-6-1-2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-467-9600

TELEFAX: 206-623-6793  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1199 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-455-602-2  
  
Query Match 3.5%; Score 220.5; DB 1; Length 1199;  
Best Local Similarity 18.6%; Pred. No. 3e-09;  
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;

Qy 357 GFYHPGVLPVNNFRRRPDQHSSTKDVSE-----AYVCLPCREG- 398  
Db 498 GTWHGVLNIDYKIQ---MNKSGMVRVSCSEPLCKGQIKVIRKGEVSCCWTACKENE 554  
Qy 399 -----ADDSPCFVQEDKYLR-----LAIISFOGLCMLDFVSMV 438  
Db 555 FVQDEFTCRACDLGWWPNAELTCEPIPVRYLEWSDIESIIAIAFSLCILVTLFVTLIF 614  
Qy 439 YHFRKAKSIRASGLLLETLFGSLLLYFPVILYFEPSTFCILLRWARLLGFATVYCT 498  
Db 615 VLYRTPVVKSSRELXYIIAGIFLGVVCPPTLIAPPTTSCYLORLLVGLSSAMCYSA 674  
Qy 499 VTLKLRHVLKV-----FLSRQAQIPYMTGGRVMRLAVILLVVFVFLIGW 544  
Db 675 LVTKNRIARILAGSKKICTRKPREFMSAWAQVIAIISVQLTLVTLII----- 726  
Qy 545 TSSVCONLEKQISLIGQKTSDLHIFNMCLIDRWDMYTAVA-EFLFLMGVYLCYAVRTV 603  
Db 727 -----MEPPMPILSYPSIKE--VYLICNTNLGVAPVGVYNGLLIMSTCTYAEKTRV 777  
Qy 604 PSARHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMYFAHTHLTVVTIGLL 663  
Db 778 FANFNEAKYIAFTMTTCTIIWLAFFVIYF-----GSNKIIITTCFAVSLSVTVALGCMF 831  
Qy 664 IPK----FSHSSNNPRDDIATEAYEDELDMG-----RSGSLNNSINSASWSEHSLDP 711  
Db 832 TPKMVIIIAKPERNVSATF---SDVVRMHVGDGKLCPCRSNTFLN----- 874  
Qy 712 EDIRDELKLYAQLEIYKRRKMITNNPHLOKKRCS-----KGLGRKIMRRITEIPT 764  
Db 875 -----IFRRKKPGAGNANSNGKSVSWSEPPGGRQAPKQGHVWQRLSVHVKT 919  
Qy 765 VSRQSKEDKADGHTAGTALIRKNPSSGNGTKSKEETLKRNVSFLKSKSHSTYDHY 824  
Db 920 NETACNQ-----TAVIK---PLTKSYQSGSKSLTFSD----- 948  
Qy 825 RDQTESSSLPTESOEETSTLESLSGKLTQKLEDSEASTESVPLVCKKSASAHN 884  
Db 949 -----ASTKTLNVEEDNTPSAHFSPSPSPSMVVR 980  
Qy 885 LSSEKKTGHPRMTSLQKSLSVIASAKEKTLGLAGKTQAGVEERTKSKQKPLPKDKETNRN 944  
Db 981 RGPVATTPLPPLHL-----TAETPLFLADSVIPKGL-----PPPLPQOQ----- 1021  
Qy 945 HNSDNTETKDPAPQNSNPAEPRKQSGIMKQORVNPPTANSDL-----NPGTTQ 996  
Db 1022 -----POQPPQPPQPK--SLMDQLQGVVTFNGSGIPDFHAYLAGPGTP- 1065  
Qy 997 MKDNFDIGEVCPWEYDITPGVPSE--SKVQKHYSIVASEMEKNPTFSLKESHHKPKA 1054  
Db 1066 -----GN-----SLRSLYPPPPPHQLMPLHLSTFQEEISIPGEDIDDDSS----- 1108  
Qy 1055 AEVCOOSNOKRIDKAEVCLWESQ---OSILEDEKLLISKTPVLPRAKEENGOPRAAN 1111  
Db 1109 -----ERFKLQEFYIEREGNTEDELEDEEDLPTASKLTPEDSPALTPPSPFRDS 1159  
Qy 1112 VCAGOSEELPP--KAVASKTENENLNQIGHQEKTSSS 1147  
Db 1160 VASGSSVPSPVSESVLCTPPNVTYASVILRDYKQSSS 1197

```

727  -----MPPMPLISPSIKE--VYLNTSNTSLGVPAPGVNGLLIMSCYYIAFKTRNV 777
604  PSAFHPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTLTVTTIGLLL 663
778  PANFNEAKYIATMYTTCIIWLAFPIYF-----GSNKKIITTCFAVSLSVTALGCMF 831
664  IPK-----FSHSSNNPRDDIATAYEDELDMG-----RSGSYLNSINSANSEHS LDP 711
832  TPKMYIIIAKPERNVRSAPFTT--SDVVRMHVGDGKLCRNSNTELN----- 874
712  EDIRDELKLYLAOLEIYKRKKMITNPHLQKRC--KGLGRSINRRITEIPET 764
875  -----IFRRKKPGAGNANSNGKSVSWSPGQRQAPKCOHVWQRLSVHVKT 919
765  VSRQCSKEDKEGADGTAKGTALIRKNPPESSGNTGKSEETLKNRVFSLKKSHSTYD 824
920  NETACNQ-----TAVIK--PLTKSYOGSKSLTFSD----- 948
825  RQOTESESSLPTESQBEETENTSTLESGLKKLTQKLKEDSEASTESVPLVCKSASAH 884
949  -----ASTKTLVNEEDNTPSAHFSPSPSPSMVVR 980
885  LSEKKTGHPRTSMLOKLSVITASAKETGLAGKTQTAGVEERTKQKPLPKDKETNRN 944
981  RGPVATTPLPPLH-----TAEETPLFLADSVIPKGL-----PPLPQOQ----- 1021
945  HNSDNTETKDPAPONSNAPEPRKPKQSGTMMKQORVNTTANSDL-----NPGTTQ 996
1022  -----POQPPQPPQPPK--SLMDQLQGVVTFNGSGIPDFHVLAVLAGPQTP- 1065
997  MKDNFDIGEVCPWEYVDLTPGPVPS--SKYOKHVSIIVASEMEKNPTFSLKEKSHHK 1054
1066  -----GN-----SLSLYPPPPQHQLQMLPLHLSTFQESISPPGEDIDDDS----- 1108
1055  AEVQOSNOKRIDKAEVCLWESQ---OSLEDEKLLISKTPTVLPERAKEENGQOPRA 1111
1109  -----ERFKLLQEFYIEREGNTEDELEEDLPTASKTLTPEDSPALTPPSPFRDS 1159
1112  VCAGOSEELPP--KAVASKTENENLNOIGHOEKKTSS 1147
1160  VASGVSPPSPSVESVCLTTPNVTYVILRDYKQSS 1197

RESULT 5
PCT-US91-09422-2
; Sequence 2, Application PC/TUS9109422
; GENERAL INFORMATION:
; APPLICANT: Mulvihill, Eileen R.
; APPLICANT: Hagen, Frederick S.
; APPLICANT: Houamed, Khaled M.
; APPLICANT: Almers, Wolfhard
; TITLE OF INVENTION: G PROTEIN-COUPLED GLUTAMATE RECEPTORS
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend
; STREET: One Market Plaza, Steuart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/09422
; FILING DATE: 19911212
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,007
; FILING DATE: 18-NAR-1991
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 07/648,481
; FILING DATE: 30-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/626,806
; FILING DATE: 12-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 467-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1199 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US91-09422-2

Query Match      3.5%; Score 220.5; DB 5; Length 1199;
Best Local Similarity 18.6%; Pred. No. 3e-09;
Matches 163; Conservative 120; Mismatches 330; Indels 265; Gaps 29;

QY 337 GFYHPGVLPVNNFRRRGPDQHSSTKDYSEE-----AVCLPCREG- 398
Db 498 GTWHEGLNDDYKIQ---MKNSMVRSCVSEPCPLKGQIKVIRKEVSCCWTACKENE 554
QY 399 -----CPFC-----ADSPCFVQEDKYLR-----LAIISFOGLCMLLDFVSMVLV 438
Db 555 FVQDEFTCRACDLGWPNAELTGEPIPVRYLEWSDIEIIAIAFSCUGILVTLFVTLIF 614
QY 439 YHFRKAKSIRASGLILLLETLFGSLLLYFPVVILYFEPSTFCILLRWALLGFATVYGT 498
Db 615 VLYRDTPVKSSRELXYILLAGIFLGVCPTLLIAKPTTSCYLQRLVGLSSAMCISA 674
QY 499 VTLKLRVLKV-----FLSRTAQRIPYMTGGVRMRLAVILLVFWFLIGW 544
Db 675 LVTKNRIARILAGSKKICKTRKPFMSAWAQVITIASLISVQLTLVVTLLI----- 726
QY 545 TSSVCNLEKOILSIGQCKTSDHLFNCLIDRDYMTAVA-EFLFLLGWVLYCYAVRTV 603
Db 727 -----MEPPMPLISYSIKE--VILICNTSLNGVAPVGYNGLLIMSCITYAFKTRV 777
QY 604 PSAPHEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTHLWTVTIGLL 663
Db 778 PANFNEAKYIAFTMTTCIWLAFVIYF-----GSNYKIITTCFAVSLSVTVALGCMF 831
QY 664 IPK-----FSSSNPNRDDIATEYEDELDMG-----RGSYLNSSINSAMSEHSLDP 711
Db 832 TPKMYIIIAKPERNVRSFTT---SDVVRMHVGDGKLCPCRSNTFLN----- 874
QY 712 EDIRDELAKLYAQLIYIRKKMTNPNHLOKKRCS-----KGLGRSIMRRIETPEI 764
Db 875 -----IFRRKAPGAGNANSKSVSWSPGGGRQAPKGQHVWVORLSVHVKT 919
QY 765 VSRQSKEDKEGADHTAGKTALIRKNPSSSGNTGKSKETLKNRVFSLKKSHTYDHY 824
Db 920 NETACNQ-----TAVIK---PLTKSYQSGSKSLTFSD----- 948
QY 825 RDQTEESSLPTESQEEETTENSTLESUGKKLTQKLEDSEASTESVPLVCKSASAHN 884
Db 949 -----ASTKTLYNVEEDNTPSAHFPSPSPSMVVRH 980
QY 885 LSSEKKTGHPRTSMLOKSLSVIASAKEITLGLAGTQTAGVEERTKSOKPLPKKETHRN 944
Db 981 RGPVATTPPLPPLH-----TAEETPLFLADSVIPKGL-----PPPLPQQO----- 1021
QY 945 HNSDNTETKDPAPONSPEEPRKPKQSGIMKQORVNPTTANSDL-----NPGPTQ 996
Db 1022 -----PQQPPQPPQPPK---SLMDQLQGVVTFNGSIGPIDPFHVLAVAGPGTP- 1065
QY 997 MKDNFDIGEVCPWEYIDLTPGVFPSE--SKVQKHVSIVASEMEKNPTFSLKEKSHHKPKA 1054
```

```

Db 1066 -----GN-----SLRSLYPPPPPPQHLQMLPLHLSTFQEEISIPPGRDIDDD- 1108
QY 1055 AEVCQSQNQKRIDKAEVCLWESQG---QSILEDEKLLLSKTPVLPERAKEENGOPRAAN 1111
Db 1109 -----ERFKILQEFVIREGNTDEDEDEEEDLPTASKLPEDSPALTPPSPFRDS 1159
QY 1112 VCAGQSEELPP--KAVASKTENENLNQIGHQEKRTSS 1147
Db 1160 VASGSSVPSSPVSESVLCTPPNVTVASVILRDYKQSSS 1197

RESULT 6
US-08-687-289A-6
; Sequence 6, Application US/08687289A
; Patent No. 5981195
; GENERAL INFORMATION:
; APPLICANT: Fuller, Forrest H.
; APPLICANT: Krapcho, Karen J.
; APPLICANT: Hammerland, Lance G.
; TITLE OF INVENTION: CHIMERIC RECEPTORS AND METHODS FOR
; IDENTIFYING COMPOUNDS ACTIVE AT
; TITLE OF INVENTION: METABOTROPIC GLUTAMATE RECEPTORS AND
; TITLE OF INVENTION: THE USE OF SUCH COMPOUNDS IN THE
; TITLE OF INVENTION: TREATMENT OF NEUROLOGICAL DISORDERS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSeq for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,289A
; FILING DATE: July 25, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,526
; FILING DATE: July 26, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-687-289A-6

Query Match      3.4%; Score 220; DB 2; Length 1219;
Best Local Similarity 18.8%; Pred. No. 3.4e-09;
Matches 174; Conservative 127; Mismatches 354; Indels 270; Gaps 34;

QY 311 QKVIDQCSSDWFSGTHKC---HLNNSE-CMPKIGLGLFVLGAYECTCKAG---FYHPGV 363
Db 475 BQVTFDEC---GDLVGNYSIINWHLSPEDGSGIVFKEVGY----YNNYAKKGERLFINEEK 527
```

```

Qy 364 LPVNNRRRRQDQH-----ISGSKDQVSE-----AYVCLPCREG-----CPFCAD 404
Db 528 ILWSGFSREVFSNCRDLGTRKGIIEGPTCCFCEVCPECDGEYSDETDASACNKP 587
Qy 405 D-----SPCFVQEDKYLR-----LAIISFOGLMCLLDVSMVLYVHFRKAKSIRAS 451
Db 588 DFWSNENHTSCEPIPVRYLEWSDIESIIAIAFSGILGILVTLFVLYRDTVPWKSS 647
Qy 452 LILLETILFGSLLYFPVILYFPEFRCILLRWARLLGFATYGVITLKLHRLVKV-- 509
Db 648 RELCYIILAGIFLGVCFEFTLIAKPTTSCYLQRLVLLGSLSAMCYSLAVTNTNRIARLA 707
Qy 510 -----FLSRTAQRIPYMTGGVRMLAVILLVFWELIGWTSVVCQNLKQIS 557
Db 708 GSKKKICTRKPFRMSAQVLIASILISVQLTLVTLII-----MEPMP 752
Qy 558 LIGOCKTSDHLIFNCLIDRWYMTAVA-EFLFLMGVLYCYAVRTVPSAFHEPRYMAVA 616
Db 753 ILSYPSIKE--VYLICNTSLNGVAVPGYNGLLIMSCYYAFKTRNVPANFNEAKYIAFT 810
Qy 617 VHNELIISAIFHTIRFVLASRLQSDWMLLYFAHTLTVTITGLLLPK-----FSHSSN 672
Db 811 MYTTCIILAFVPIYF-----GSNKIITFCFVSLSVTVVALGCMFTPKMYIIIAKPER 864
Qy 673 NPRODIAFEAYEDELDMG-----RSGSYLNSSINSAMSEHSLDPEDIRDELKLYAQ 724
Db 865 NVRAFTT--SDVVRHVHGKLPKRSNTFLN-----894
Qy 725 LEIYKRKKMITNPHLOKKRCS-----KKGLGRSIRMRRIETVTSROCKEDKEGA 777
Db 895 --IFRRKKPGAGNANSKGSVSEPGGRQAPKGQHVQWRLSVHVKTNETACNQ-----946
Qy 778 DHGTAKTALRKPPSSNGTNGSKETLKNRVFSLKKSHTYDHRVDQTEESSSLPTE 837
Db 947 -----TAVIK-----PLTKSYQSGKSLTFS-----968
Qy 838 SQEBETTENSTLESIGSKKLTOKLKEDSEASTESVPLVCKSAHNLSEKKTGHPRTS 897
Db 969 -----ASTKTLYNVEEDNTPSAHFSPSSFMVVRHGGPPVATTPPLP 1013
Qy 898 MLQKSLSVIASAKETLGLAGKTOTAGVEERTKSKQLPKDKETNRNHSNDTETKDP 957
Db 1014 HL-----TAEETPLFLADSVIPKGL-----PPLPQQ-----1041
Qy 958 PONSNAPEERPKQSGIMKQOQVNPNTANSDL-----NPGTTQMKDNFDIGECPW 1009
Db 1042 PQPPPPQPPQPK--SLMDQLQGVTVNFGSGIPDFHVLVLAGPTP-----GN-----1087
Qy 1010 EVDLTGPPVPE--SKVQKHVSIVASEMEKNPTFSLKEKSHHKPKAAEVCQSQNKRID 1067
Db 1088 SLRSLYPPPPPHQLMPLHLSTFQESISPPGEDIDDDDS-----EREF 1132
Qy 1068 KAEVCLMESQ---OSILEDEKLILSKTPVLPERAKENGQPPRAANVCAGQSLELPP-- 1122
Db 1133 LLOEFVVEREGTEDELEEDLPTASKLTPEDSPALTPSPPRDSVASGSSVPSPVPS 1192
Qy 1123 KAVASKTENENLIGHQEKTKSS 1147
Db 1193 ESVLCTPPNTYASVILRDYKQSS 1217

```

RESULT 7

```

US-08-660-148-2
; Sequence 2, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,148
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-9419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-660-148-2

Query Match 3.2%; Score 204; DB 4; Length 1180;
Best Local Similarity 19.0%; Pred. No. 7.le-08;
Matches 178; Conservative 120; Mismatches 401; Indels 240; Gaps 34;

Qy 219 ETEFHLGRKWRPHLRRGPNQGRGLGHSWR-----RKDGLGGDKSHKWSPPYLECE 273
Db 333 DVKWFDDYILKLRPETNHRNP-----WFQEFWHRFCRLEGPEQSESKYNT-----CN 382
Qy 274 NG-----SVKPGMLV-TLSSAIYGL---QPNLVPEFRGV-----MKVDINLQ 311
Db 383 SSUTLTKTHHVQDSKMGVFNIAIYSMAYLHNMQMSLCPGYAGLCDAMKPIDGRKLLESML 442
Qy 312 KVIDOCSSDGFSGTHKCHLNNESECPIKGLGVFLGAYECICKAGFYHGPVLPVNNFR 371
Db 443 KTNFTGVSGDTILFDENGSDPGRYEIMNFEMG-----KDYFDYINVGSWDNGEL 492
Qy 372 RGPDOHI---SGSTKDVSEE-----AYVCLPCREG-----CPFC 402
Db 493 KMDDEYMSKSNIRSVCEPCEKGOIKVIRKEVSCCWTCTPCKENYVDFDEYTCAC 552
Qy 403 -----ADD-SPCFVQEDKYLR-----LAIISFOGLMCLLDVSMVLYVHFRKAKSIRA 449
Db 553 QLSGWSPTDDLTGCOLIPVQYLRMGDPEPIAAVFPACGLIATLFTVTVFVIYRDTVPVKS 612
Qy 450 SGLILETILFGSLLYFPVILYFPEFRCILLRWARLLGFATYGVITLKLHRLVKV 509
Db 613 SSRELCVIILAGICGLYCLCTFCLIAKPKQIYCVLQRIIGISLSPAMSYSALVTNTRIARI 672
Qy 510 -----FLSRTAQRIPYMTGGVRMLAVILLVFWFLIGWTSVVCQNLKQ 555
Db 673 LAGSKKKICTRKPFRMSAQV-----LVIAFIL-----CIQIGII 708
Qy 556 ISL-----IGOKTSDHLIFNCLIDRWYMTAVA-EFLFLMGVLYCYAVRTVPSAF 607
Db 709 VALFIMEPPDIIHMDYPSIREVYILCNITNLTGVTPGLYNGLLILSCTFYAFKTRNVPANF 768
Qy 608 HEPYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLLYFAHTLTVTITGLLLPK- 666
Db 769 NEARYIAFTWYTCIIWLAFVPIYF-----GSNKIITMCFVSLSATVALGCMFVPKV 822

```





RESULT 9  
US-08-367-264-8  
; Sequence 8, Application US/08367264  
; Patent No. 6001581  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen  
; APPLICANT: Pontsler, Aaron  
; APPLICANT: Johnson, Edwin C.  
; APPLICANT: Hess, Stephen D.  
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,  
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/367,264  
; FILING DATE: 02-JUN-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/072,574  
; FILING DATE: 04-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: FP41 9772  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4737  
; TELEFAX: 619-546-9392  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1180 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; 08-367-264-8

Query Match 3.1%; Score 200; DB 3; Length 1180;  
Best Local Similarity 18.9%; Pred. No. 1.6e-07;  
Matches 178; Conservative 116; Mismatches 401; Indels 248; Gaps 33;

QY 219 ETEWFGHRRKWRPHLRGPNQGRGLGHSWRRKDLGLGDKSHFKWSPPPYCEGNGSY- 277  
DB 333 DVKWFDDVYLKRPETNRNP-----WFQEFWQHR-----FQCRLEAFPQENSKYN 378  
QY 278 -----KPGWLVL-TLSSAIYGL---QPNLVPEFRGV-----MKVD 307  
DB 379 KTCNSSLTKTHVQDSKMGFVINAIYSWAGLHNMQLCPGYAGLGDAMKPIDGRKLL 438  
QY 308 INLQKVDIDQSSDGFSGTHKCHLNNECMPKIGLGFVLGAYECICKAGFYHPCVLPVN 367  
DB 439 ESLMTNTGTGSGDILFDENGSDSPRIEMNFKMG-----KDYFDYINVGSD 488  
QY 368 NFRRRGPQHI-----SGSTKDVSEE-----AYVCLPREG----- 398  
DB 489 NGELKMDDEVMSKSNIRSVSCPECKGQIKVIRKGEVSCCTCTPCKENEYVFDVYT 548  
QY 399 CPFC-----ADD-SPCFVQEDKYL-----LAIISFGGLCMLLDFVSMVLVYHFRKAK 445  
DB 549 CKACQLSGWPTDGLTGPVQLRWGDPPEIAAVVFACILGLATLFTVTVVFIYRTP 608

QY 446 SIRASGLILETLFGSLLLYFPVVLVYFEPSTFRCILLRWARLLGFATVYGTVLKLR 505  
DB 609 VVKSSRELICYIILAGICGLYLTCTCLIAKPKQICYIQRIGIGLSPAMSYALVTKNR 668  
QY 506 VLKV-----FLSRTAQRIPYMTGGRVMRLAVILLVYFVFLIGHTSVVCON 551  
DB 669 IARILAGSKKICKTPKPRFMSACAQ-----LVIAFIL-----CIQ 704  
QY 552 LEKQISL-----ICQKTSDLHIFNMCLIDRWDMYMTAVA-EFLFLLGWVYLCVAVRTV 603  
DB 705 LGIIVAFIMEPPDIMHDYPSIREVYLICNTNLGVVTPGLNNGLLILSCTFYAEKTRN 764  
QY 604 PSARHEPYMAVAVHNEILLISAFHTIRFVLASRLQSDWMLMLYFAHTHLTVTVTIGLL 663  
DB 765 PANFPEAKYIAFTTTCIIWLAFVIYF-----GSNKIITMCFVSLSLTVALGCMF 818  
QY 664 IPK-----FSSNSNPRDDIATE-AYEDELDMGRSGSYL--NSSINSAMSEHSLDPEIRD 716  
DB 819 VPKVYIILAKPERNVRSFTTSTVYRMHVGDGCKSSAASRSSLYNLNKKRRSGSSETLSS 878  
QY 717 ELKK-----LYAQLEIYKRRKMITNPNPHLOK---KRCSSKGLGRSMTMRIT 759  
DB 879 NGKSVTWAQNEKSSRGQHLWQRLSIHNKENPNCTAVIKPPKSTESRGLAG----- 932  
QY 760 EIPETVSRQCKEDKEGADHGTAKGTALIRKNPPSSNGTGSKEETLKNRVESLKKSHS 819  
DB 933 -----AGAG-GSAGGVGATGGAGAGAGGPGSPDAGPKALY----- 969  
QY 820 TYDVRDQTESSSLPTSEEBEETTENTLESLSKKTQKLKEDSEABESTESVPLVCKS 879  
DB 970 -----DVAEAEHPAPARPSPISTLSHRAG---SASRTDDVPSLHSEPVARS 1019  
QY 880 ASAHNLSSEKKTGHPR-TSMLOKLSLVITASAKEKTLGLAG-----KTQTA 923  
DB 1020 SQSGLEQIISVTVTRFTANISELNMMLSTAAPSGVGAPLCSSYLIPKEIQLPTTMTT 1079  
QY 924 GVEERTKQKPLPKDKETNRNHSNDTETKDPAPQN--SNPAEPRKPKQSGIMKQORV 981  
DB 1080 FAE-----IQPLPAIEVTGGGAQAQAAGAAADARESFAAGPEAAAKPDLEELVALTPP 1134  
QY 982 NPTTANSDLNPGTOMKONFDIGEVCPWEVVDLTGPGVPVSESK 1024  
DB 1135 SP-----FRDSVDSGSTTPNSPVSESAICIPSPSPK 1164

RESULT 10  
US-09-153-757-8  
; Sequence 8, Application US/09153757  
; Patent No. 6413764  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie  
; Ellis, Steven B.  
; Liaw, Chen  
; Pontsler, Aaron  
; Johnson, Edwin C.  
; Hess, Stephen D.  
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,  
; NUCLEIC ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/153,757

```

; FILING DATE: 15-Sep-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: FP41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1180 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
09-153-757-8
Query Match 3.1%; Score 200; DB 4; Length 1180;
Best Local Similarity 18.9%; Pred. No. 1.6e-07;
Matches 178; Conservative 116; Mismatches 401; Indels 248; Gaps 33;
QY 219 ETEWFHGLRRKWRPHLHRRGPNQGRGLGHSWRRKDGGLGDKSHFKWSPPPYLECENGSY- 277
DB 333 DVKWFDDYLLKLRPTNHRNP-----WQEFWQHR-----FQCKLEAFPQENSYN 378
QY 278 -----KPGMLV-TLSSAIXGL---QPNLVPEFRGV-----MKVD 307
DB 379 KTCNSSLTKTHHVQDSKMGFVINAISYMAVGLHNMQSLCPGYAGLCDAMKPIDGRKLL 438
QY 308 INLQKVDIDQSSDGFSGTHKCHLNNECPHKGILGVLCAYECICKAGHYHFGVLPVN 367
DB 439 ESLMKTNTGTSGDTILFDENGSDSPGRYEIMNFKEMG-----KDYFDYINVGSWD 488
QY 368 NFRRRGPDQHI-----SGSTKDVSE-----AYVCLPCREG----- 398
DB 489 NGELKMDDEVWKSNSNIIRSVSPCEKQIKVIRKEVSCCWTCTCKENEYVFDEYT 548
QY 399 CPFC-----ADD-SPCFVQEDKYL-----LAIISFQGLCMLDFFVSLVYHFRKAK 445
DB 549 CKACGLGSWPTDGLTGCDDLPVQYLRWGDPEPIAAVFAECLGLLATLFVTVFIIYRDP 608
QY 446 SIRASGLILLETILPGSLLLYFPVILYFEPSTFCILLRWARLLGFATVYGVTLKLR 505
DB 609 VYKSSRELCTIILAGICIGVLCFTCLIAKPKQIYCIYQIRIGLSPAMSYSAIVTKNR 668
QY 506 VLKV-----FLSRTAQRIPYMTGGRVWRMLAVILLVWFVFLIGWTSVVCQN 551
DB 669 IARIILAGSKKIKCTPKPFMSACAO-----LVIAEILI-----CIQ 704
QY 552 LEKQISL-----IGOKTSDHLFNNCLIDRWDMYTAFA-EFLFLMGVLYCVAVRTV 603
DB 705 LGLIIVAFIMEPPDLMHDYPSIREVYLICNTNLGVVTPPLGNGLLILSCTFYAFKTRV 764
QY 604 PSAPHEPRYMAVAVNELIISAIFHTIRFVLASRLOSDDMLMLYFAHTHLVTVTIGLL 663
DB 765 PANFPEAKYIAFTYTTTCIIWLAFVPIYF-----GSNYKIITCMCFVSLSATVALGCMF 818
QY 664 IPK-----FSHSSNNPRDDIATE-AYEDELDMGRSGYL--NSSINSAMSEHSLDPEDIRD 716
DB 819 VPKVYIILAKPERNVRSAFTSTVVRMHVGDGKSSAASRSSSLVNLWKRSGSGELUSS 878
QY 717 ELKK-----LYAQLEIYKRRKMLTNNPHLQK---KRCSSKKGLGRSITMRRIT 759
DB 879 NGKSVYTAQNKSSRGQHLWQRLSHINKENPNQTAIVIKFPKSTESRGIGAG----- 932
QY 760 EIPETVSRQCSKEDKEGADHCTAGKTALIRKNPPSSGNTCKSKEETLKNRVSLKSHS 819

```

```

DB 933 -----AGAG-GSAGGVGATGGAGCAGAGPGGPGSPDAGPKALY----- 969
QY 820 TYHVRDQTESSSLPTESQEEETENSTLESGLKTKLKEDSEAESVPLVCKS 879
DB 970 -----DYAEAEHFPAPAPRSPSPISTLSHRAG-----SASRTDDDDVPSLHSEPVARS 1019
QY 880 ASAHNLSEKKTGHPR-TSMLQKSLSVTASAKETKLAG-----KTQTA 923
DB 1020 SSQGLMEQISSVTVTRFTANISLNSMMLSTAAPSPGVGAPLCSYLLIPKEIQLPTTMTT 1079
QY 924 GVEERTKSQKPLPKDKETNRNHSNDNTETKDPAPQN--SNPAEPRKPKQSGIMKQORV 981
DB 1080 FAE-----IQPLPALEYTGGAQPAAGAAQDAARSPAAAGPEAAAKPDLEELVALTPP 1134
QY 982 NPTTANSDLNPGTTQMKDNFDIGVCPWEVYDLPFGPVVPSESK 1024
DB 1135 SP-----FRDSVDSGSTTPNSPVSESALCIPSSPK 1164

RESULT 11
US-08-660-148-5
; Sequence 5, Application US/086600148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,148
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-9419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1212 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-660-148-5

Query Match 3.1%; Score 197.5; DB 4; Length 1212;
Best Local Similarity 18.6%; Pred. No. 2.6e-07;
Matches 207; Conservative 135; Mismatches 363; Indels 387; Gaps 48;
QY 219 ETEWFHGLRRKWRPHLHRRGPNQGRGLGHSWR-----RKDGLGDKSHFKWSPPYLECE 273
DB 333 DVKWFDDYLLKLRPTNHRNP-----WQEFWQHRFQCRLEGFPQENSYKNT-----CN 382

```

QY 274 NG-----SYKPGWLVTLSAIYGL---QPNLVPFRGV-----MKVDINLQ 311  
Db 383 SSLTLKTHHVQDSKMGFVINALYSMAYGHLNMQMSLCFGYAGLCDAMKPIDGRKLLESIM 442  
QY 312 KVIDQCSDDGWFSTHKKHNLNSECMPKGLGVLGAYECICRAGFYHPGVLPVNNFR 371  
Db 443 KTNFTGVSGDITLFDENGSPGRVEIMNFEMG-----KDYFDYINVGSMDNCEL 492  
QY 372 RGPQDHI-----SGSKVDYSEE-----AYVCLPCREG-----CPFC 402  
Db 493 KMDDDEVWKSNIIRSVCSPECEKQKIVIRKGEVSCCWTCTPCKENEYVDFEYTKAC 552  
QY 403 -----ADD-SPCFVQEDKYLR-----LAIISFOGLCMLLDFVSMVYVHFRKAKSIRA 449  
Db 553 QLGSNPTDGLTGCDLIPQYLWGDPEPIAAVFAVACIGLLATLFTVVVFIYRDPVVK 612  
QY 450 SGLIILETLFGSLLLYPVVPVILYFETFRFCILLRMARLLGFATVYGTVLKHLRVLKV 509  
Db 613 SSRELCYIILAGICGLVCTCLIAKPKQIYCYLQIGIGLSPAMSYSALVTKTNRARI 672  
QY 510 -----FLSRTAQRIPIYMTGRVNRMLAVILLVFWFLIGWTSVQCNLEKQ 555  
Db 673 LAGSKKIKCTKKPRMSACQ-----LVIAETLI-----CIGLGI 708  
QY 556 ISL-----ICQKQTSDBLIFNMCLIDRWDMYTAVA-EFLFLWGVVLYCVAVRTVPSAF 607  
Db 709 VALFTIMEPPDIDHDPYSIREVYLICNTNLGVTPGLVNGLLILSCFYAEKTRVNPANF 768  
QY 608 HEPYMAVINELIISAIFHTIRFVLASRLQSDMMLYFAHHTLTVTVTIGLLIPK- 666  
Db 769 NEAKYIAFTMTTCIILAFPIVF-----GSNYKIITMFSVLSATVALGCMFVPKV 822  
QY 667 ---FSHSSNNRRDDIATE-AVEDELMGRSGSYL--NSSINSANSEHSLDPEDIRDELKK 720  
Db 823 YIIIAKPERNVRSFATFTSVYVMHVGDKSAAASRSLSNLNLMKRRGSGSETLR----- 877  
QY 721 LYAQLIEYKRRKMTNPHLOKRCCKKGLGRSIRMRRITEIPETVSRQCSKEDKEGADHG 780  
Db 878 -----YKDRRLAQHKEI-----ECF----- 893  
QY 781 TAKGTALIRKNPPSGNGTGSKEETLKNRVFLSKKSHSYDHRVDQTEESSSLPTESQ 840  
Db 894 TPKG-----SMNGGGA-----TWSSNGKSVTWAQN 920  
QY 841 EETENSTLESGLKKTOKL-----KEDSE-----AESTESVPLVCKSASAHNLS 886  
Db 921 EKSR-----GOHLWQRLSIHINKENPNQTAIVIKPPFKSTESRGLGAGAGGSAG 972  
QY 887 SEKKTG-----HPRTSMLQKSLSVIASAKEKTLGLAGKTQTAGVEERTKSQK 933  
Db 973 GVGATGGACAGAGGGPESPDAG--PKALYDVAEAEHFPAPA-----RPRSPS 1020  
QY 934 PLPKDKETNRNHSNDTETKDPAPQ-NSNPAERPRPKQSGIMKQ--QRVNPPTAN-SD 989  
Db 1021 PI-----STLSHRAGSARTDDVPSLHSEPVARSSSQGS-LMEQISSVVVTRFTANISE 1074  
QY 990 LNPCTTQMKDNFDIGEVCPWEYDITFG---PVPSESKVQKHVSIVASEMEKNFTFSLKE 1046  
Db 1075 LN-----SMULSTAAP-----SPGVGAPLCSVYLIPKEIQL-----PT----- 1107  
QY 1047 KSHHKPRAAEVCOQSNQKRIDKAEVCLWESQGSQSILEDKLLISKITPVLPERAKEENG- 1105  
Db 1108 -----TWITFAEQPLPAIEVTGGA 1127  
QY 1106 OPRAANVAC-----OSEELPPKAVASKTENENLNOIGHQEKKTSSSEENVRSYNSNNF 1161  
Db 1128 QPAAGAQAAGDAARESPAAGPEAAAKPDLEELVAL-----TPSPFRDSVDSGST- 1178  
QY 1162 QOPLTSRAEVCPEFEETPAQPNAGRVALPAS 1193  
Db 1179 -----TPNSVPSSEALCIPSS 1194

RESULT 12  
US-08-486-270-10  
; Sequence 10, Application US/08486270  
; Patent No. 5807689  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen  
; APPLICANT: Pontsler, Aaron  
; APPLICANT: Johnson, Edwin C.  
; APPLICANT: Hess, Stephen D.  
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,  
; NUCLEIC ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/486, 270  
; APPLICATION NUMBER: US/08/486, 270  
; FILING DATE: 02-JUN-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/072,574  
; FILING DATE: 04-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: FP41 9772  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4737  
; TELEFAX: 619-546-9392  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1212 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-486-270-10

Query Match 3.0%; Score 193.5; DB 1; Length 1212;  
Best Local Similarity 18.5%; Pred. No. 5.7e-07;  
Matches 207; Conservative 131; Mismatches 383; Indels 395; Gaps 47;  
QY 219 ETEWFGHLRRKWRPHLHRRGPNQGRGLGHSHRRKGLGGDKSHFKWSPPYLEGENSY- 277  
Db 333 DVKWFDDYLLKRLRPTNHRNP-----WFQEFWQHR-----FQCRLEAFPOENSKYN 378  
QY 278 -----KPGWLVTLSAIYGL---OPNLVPFRGV-----MKVD 307  
Db 379 KTCNSSLTKTHHVQDSKMGFVINALYSMAYGHLNMQMSLCFGYAGLCDAMKPIDGRKLL 438  
QY 308 INLOKVVIDQCSDDGWFSTHKKHNLNSECMPKGLGVLGAYECICRAGFYHPGVLPVN 367  
Db 439 ESLMKTNFTGVSGDITLFDENGSPGRVEIMNFEMG-----KDYFDYINVGSMD 488  
QY 368 NFRRRGPDQHI-----SGSTKDVSEE-----AYVCLPCREG----- 398  
Db 489 NGELKMDDDEVWKSNIIRSVCSPECEKQKIVIRKGEVSCCWTCTPCKENEYVDFEY 548  
QY 399 CPFC-----ADD-SPCFVQEDKYLR-----LAIISFOGLCMLLDFVSMVYVHFRKAK 445  
Db 549 CKACQLGSWPTDGLTGCDLIPQYLWGDPEPIAAVFAVACIGLLATLFTVVVFIYRDP 608  
QY 446 SIRASGLILLETILFGSLLLVFPVVIYFEPSTPRCILLRWARLLGFATVGTVTCLKHR 505

Db 609 VVSSRELCTIILAGICIGYLTCTFCLAKPKQIYQLORIGIGLSPAMSYALVTKTNR 668  
QY 506 VLKV-----FLSRTAQRIPTYMGGRVRLAVILLVVFVFLIGWTSVVCQN 551  
Db 669 IARILAGSKKICTPKPREMSACQ-----LVIAFILI-----C1Q 704  
QY 552 LEKQISL-----IGQKTSDBLIFNMLIDRWDMYATA-EFLLMGVYLCYAVRTV 603  
Db 705 LGIIVAFIMEPPDIMHDYPSIREVYLICNTNLGVVPLGNNGLLILSCTFYAFKTRV 764  
QY 604 PSAPHEPRYMAVAVHNELIISAFITIRVFLASRQSDMMLMYFAHTHLTVTVTIGLL 663  
Db 765 PANFPEAKYIAFTMTTCIWLAFVPIYF-----GSNKIITMCFVSLSATVALGCMF 818  
QY 664 IPK-----FSHSSNPRDDIATE-AYEDELDMGRSGYL--NSSINSAWSEHSLDPEDIRD 716  
Db 819 VPKVYIILAKPERNVRASATTTSTVVMHVGDKSSAARSLSLWMLKRRGSSGETLR- 877  
QY 717 ELKLYAQLLEIYKRRKMITNPHLQKRCRKGLGRSIMRRITEIPETVSROCSEKDEK 776  
878 -----YKDRRLAQHKSEI-----ECF----- 893  
QY 777 ADHGTAQKALIRKNPSSSGNTGSKETLNKRVFLSKKSHSTVDHVRDQTEESSSLPT 836  
Db 894 -----TPKG-----SMNGNGRA-----TMSSNGKSVT 916  
QY 837 ESQEEETTENSTLESIGKKTQKL-----KEDSE-----AESTSVPLVCKSASA 882  
Db 917 WQONEKSSR-----GOHLWORSIHINKENPNQTAIVKPPKSTESRGLGAGAGAG 968  
QY 883 HNLSEKKTG-----HPRTSMLOKSLSVIASAKETLGLAGKTQTAGVEERT 929  
Db 969 GSAGGVGATGAGCAGACAGGPGESPAG--PKALYDVAEAEHFPAPA-----RP 1016  
QY 930 KQOKPLPKDKEYNRNHNSDNTETDPAQ--NSNPAEPRKPKQSKIMKQ--QRVNPPTA 986  
Db 1017 RSPSPIT-----STLSHRAGSASRTDDVPSLHSEPVARSSSQGS-LMEQIISVVTRFTA 1070  
QY 987 N-SDLNPGTQTMKDNFDIGEVCPWEYDITPG--PVPSESKVQKHVSIVASEMEKNPTF 1042  
Db 1071 NISELN-----SMMLSTAP-----SPGVAPLCSYILPKRIQL-----PT- 1107  
QY 1043 SLKEKSHHKPAAEVQCSQSNQKRIKAEVCLMESQSQSILEDEKLISKTPVLPRAKEE 1102  
Db 1108 -----TMTFAEIQLPAIEV 1123  
QY 1103 NGG-QPRAANVCAG-----QSEELPKAVASKYTENENLMOIGHOEKKTSSSEENVRGSYNS 1157  
Db 1124 TGAQAQAQAAGDAARESPAAGPEAAAKPDLELVAL-----TPSPFPRDSVDS 1175  
QY 1158 SNNFOPLTSRAEVCWPEFETPAQPNAGRSVALPAS 1193  
Db 1176 GST-----TPNSPVSEALCIPSS 1194

RESULT 13  
US-08-367-264-10  
Sequence 10, Application US/08367264  
Patent No. 6001581  
GENERAL INFORMATION:  
APPLICANT: Daggett, Lorrie  
APPLICANT: Ellis, Steven B.  
APPLICANT: Liaw, Chen  
APPLICANT: Pontsler, Aaron  
APPLICANT: Johnson, Edwin C.  
APPLICANT: Hess, Stephen D.  
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,  
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/367,264  
FILING DATE: 02-JUN-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/072,574  
FILING DATE: 04-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Reiter, Stephen E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: FP41 9772  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-4737  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1212 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-367-264-10  
Query Match 3.0%; Score 193.5; DB 3; Length 1212;  
Best Local Similarity 18.5%; Pred No. 5,7e-07;  
Matches 207; Conservative 131; Mismatches 383; Indels 395; Gaps 47;

QY 219 ETEWFHGLRRKWRPHLHRRGPNQGRPLGHSHRRKDGGLGDKSHFKWSPPYLECENGSY- 277  
Db 333 DVKWFDYDKLRLPETNHRNP-----WQEFWQHR-----FQCRLEAFQENSKYN 378  
QY 278 -----KPNLV-FLSSAIYGL---QPNLVPEFRGV-----MKVD 307  
Db 379 KTCNSSLTKTHVQDSKMGFVINAIYSMAYGLHNNQMSLCPCYAGLCLDAMKPIDGRKLL 438  
QY 308 INLQKVIDQCSDDGWFSGTHKCHLNNSCMPTIKGLFVLGAYECICKAGFYHPGVLPVN 367  
Db 439 ESLMKTNTFTGVSGDTILFDENGDSGRYEIMNFKMG-----KDYFDYINVGSD 488  
QY 368 NFRRRGPDQHI-----SGSTKDVSEE-----AYVCLPREG----- 398  
Db 489 NGELKMDDDDEVMSKKSNIIRSVCSPECKGQIKVIRKGEVSCCWTCTPCKENEYVFEYDT 548  
QY 399 CPFC-----ADD-SPCFVQEDKYL- ----LAISFOGLCMLLDFVSLVYVYHFKAK 445  
Db 549 CRACQLGWSWPTDDLTGCDLIPVOYLRLWGDPEPIAAVVFACLGLLATLFTVTVFIYRDP 608  
QY 446 SIRASGLILETLILFGSLLLYFPVVILYFEPSTFRCILLRWARLLGFATVYGVTVTLKLRH 505  
Db 609 VVKSSRELCTIILAGICIGYLTCTFCLAKPKQIYQLORIGIGLSPAMSYALVTKTNR 668  
QY 506 VLKV-----FLSRTAQRIPTYMGGRVRLAVILLVVFVFLIGWTSVVCQN 551  
Db 669 IARILAGSKKICTPKPREMSACQ-----LVIAFILI-----C1Q 704  
QY 552 LEKQISL-----IGQKTSDBLIFNMLIDRWDMYATA-EFLLMGVYLCYAVRTV 603  
Db 705 LGIIVAFIMEPPDIMHDYPSIREVYLICNTNLGVVPLGNNGLLILSCTFYAFKTRV 764  
QY 604 PSAPHEPRYMAVAVHNELIISAFITIRVFLASRQSDMMLMYFAHTHLTVTVTIGLL 663  
Db 765 PANFPEAKYIAFTMTTCIWLAFVPIYF-----GSNKIITMCFVSLSATVALGCMF 818  
QY 664 IPK-----FSHSSNPRDDIATE-AYEDELDMGRSGYL--NSSINSAWSEHSLDPEDIRD 716

Db 819 VPKVYIILAKPERNVRSAFTTSTVVRMHVGDGKSSAASSSLVNLKRRGSSGETLR- 877  
QY 717 ELKKLYAQLEIYKRRKMTNNPHLOKKRCSKKGLGRSIMRRITEIPETVSRCCSKEDKEG 776  
Db 878 -----YKDRLAQHKEI-----ECF----- 893  
QY 777 ADHGTAQTALIRKNPPSSGNTGSKKEETLKNRVFSLKSHSTYDHRDQTESSSLPT 836  
Db 894 ----TPKG-----SMNGGRA-----TMSSNGKSVT 916  
QY 837 ESQEBETTENSTLESLSGKKLTOKL-----KEDSE-----AESTESVPLVCKSASA 882  
Db 917 WAQNEKSSR-----GOHLWQLRSLIHINKKENPNQTAIVKPFKSTESRGLGAGAG 968  
QY 883 HNLSSSEKKTG-----HPRTSMLOKSLSVIASAKETLGLAGKTTOTAGVEERT 929  
Db 969 GSAGGVATGGAGCAGAGGPGESPAG--PRALYDVAEEHFPAPA-----RP 1016  
QY 930 KSQKPLPKDKETNRNHSNDNTETKDPAPQ-NSNPAEPRKPKQSGIMKQ--QRVNPTTA 986  
Db 1017 RSPSPI-----STLSHRAGSASRTDDVPSLHSEPVARSSSQGS-LMEQISSVVTRFTA 1070  
QY 987 N-SDLNPGTTQMKDNFDIGEVCPWEYDITPG---PVPSESKVQKHVSIVASEMEKNPTF 1042  
Db 1071 NISELN-----SMMLSTAAP-----SPGVGAPLCSYLIPEIQL-----PT- 1107  
QY 1043 SLKEKSHHKPKAAEVCOQSNOKRIDKAEVCLWESQOSILEDEKLLISKTPVLPERAKEE 1102  
Db 1108 -----TMTFAEIQPLPAIEV 1123  
QY 1103 NGG-QPRAANVOCAG---QSEELPPKAVASKTENENLNOIGHQEKTSSEENVRGSSYNS 1157  
Db 1124 TGGAPQAAQAQAAGAAARESPAAGPEAAAKPDLEELVAL-----TPPSPFRDVS 1175  
QY 1158 SNNFOPLTSRAEVPWEPTETAQPNAGRSVALPAS 1193  
Db 1176 GST-----TPNSPVSEALCIPSS 1194

RESULT 14  
US-09-153-757-10  
; Sequence 10, Application US/09153757  
; Patent No. 6413764  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie  
; Ellis, Steven B.  
; Liaw, Chen  
; Pontsler, Aaron  
; Johnson, Edwin C.  
; Hess, Stephen D.  
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS.  
; NUCLEIC ACIDS ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
; STREET: 444 South Flower Street, Suite 2000  
; CITY: Los Angeles  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/153,757  
; FILING DATE: 15-Sep-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,270  
; FILING DATE: 02-JUN-1994  
; APPLICATION NUMBER: US 08/072,574

; FILING DATE: 04-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reiter, Stephen E.  
; REGISTRATION NUMBER: 31,192  
; REFERENCE/DOCKET NUMBER: FPA1 9772  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-546-4737  
; TELEFAX: 619-546-9392  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1212 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-09-153-757-10

Query Match 3.0%; Score 193.5; DB 4; Length 1212;  
Best Local Similarity 18.5%; Pred. No. 5,7e-07;  
Matches 207; Conservative 131; Mismatches 383; Indels 395; Gaps 47;

QY 219 ETEWFHGLRRKWRPHLHRRGPNQGRGLGHSWRRKDGGLGGDKSHFKWSPYPYLECENGSY- 277  
Db 333 DYKWFDDYLLKRPETNHRNP-----WFQEFWQHR-----FQCRLEAFPOENSRYN 378  
QY 278 -----KPGMLV-TLSSAIYGL---QPNLVPEFRGV-----MKVD 307  
Db 379 KTCNSSLTLKTHHVODSKMGFVINAIYSWAYGLHNNQWSLCPYAGLCDDAMKPIDGRKLL 438  
QY 308 INLQKVIDQCCSSDGFSGTHKCHLNNSCEMPIKGLGFLVGLAYECICKAGFYHVGVLV 367  
Db 439 ESLMTNFTGVSQDITLFDENGDSGRVEIMNFKEWG-----KDYFDYINVGSMD 488  
QY 368 NFRRGPPQHI-----SGSTKDVSEE-----AYVCLPCREG----- 398  
Db 489 NGELKMDDEWWSKSNIRSVCEPCEKQTKVIRKGEVSCCWTCTPCKENEYVFDEYT 548  
QY 399 CPFC-----ADD-SPCFVQEDKYLK-----LAISFOGLCMLLDFVSMVYVYHFRKAK 445  
Db 549 CKACQLGWSWPTDDLTGCDLIPVQYLRWGDPEPIAAVAVFACLGLLATLFTVTVFIYRDT 608  
QY 446 SIRASGLILETILFGLSLLLYFPVVILYFEPSTFCILLRWALLGFATVYGTVTLKLRH 505  
Db 609 VVKSSRELCTIILAGICLGYLCTFCLIAKPKQIYCYLQIRIGLSPAMYSALVTKTNR 668  
QY 506 VLKV-----FLSRTAQRIDPYMTGGRVWRMLAVILLVVFVFLIGWTSVVCQN 551  
Db 669 IARILAGSKKICTPKPRFMSACQ-----LVIAFLI-----CIQ 704  
QY 552 LEKQISL-----IGQKTSDHLLIFNMCLIDRWDMYTAVA-EFLFLWGLVYLCYAVRTV 603  
Db 705 LGIIVALIMEPPDILMHDPYPSIREVYLICNTNLGVVTPLGNNGLLILSCIFYAKTRNV 764  
QY 604 PSAPHEPRYMAVAVHNELIIISAIFHTIRFVLASRLQSDWMLMLYFAHTHLFTVTVTIGLL 663  
Db 765 PANPEEAKYIAFTMTTCIIWLAFVPIVF-----GSNKIITMCFVSLSATVALGCMF 818  
QY 664 IPK-----FSSHNNPRDDIATE-AYEDELDMCRSGSYL--NSSINSANSEHSLDPEDIRD 716  
Db 819 VPKVYIILAKPERNVRSAFTTSTVVRMHVGDGKSSAASSSLVNLKRRGSSGETLR- 877  
QY 717 ELKKLYAQLEIYKRRKMTNNPHLOKKRCSKKGLGRSIMRRITEIPETVSRCCSKEDKEG 776  
Db 878 -----YKDRLAQHKEI-----ECF----- 893  
QY 777 ADHGTAQTALIRKNPPSSGNTGSKKEETLKNRVFSLKSHSTYDHRDQTESSSLPT 836  
Db 894 ----TPKG-----SMNGGRA-----TMSSNGKSVT 916  
QY 837 ESQEBETTENSTLESLSGKKLTOKL-----KEDSE-----AESTESVPLVCKSASA 882  
Db 917 WAQNEKSSR-----GOHLWQLRSLIHINKKENPNQTAIVKPFKSTESRGLGAGAG 968









```
QY 241 QGPRGLGHSWRKDGKSHFKWSPPYLECENGSYKPGWLVTLSAIYGLQPNLVPEF 300
Db 241 QGPRGLGHSWRKDGKSHFKWSPPYLECENGSYKPGWLVTLSAIYGLQPNLVPEF 300
QY 301 RGVMKVDINLQKVDIDQCSSDGFSGTHKCHLNNSCMPKIGLGFVLGAYECICAGFYH 360
Db 301 RGVMKVDINLQKVDIDQCSSDGFSGTHKCHLNNSCMPKIGLGFVLGAYECICAGFYH 360
QY 361 PGVLPVNNFRRRGPDQHIISGSKDVSEAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420
Db 361 PGVLPVNNFRRRGPDQHIISGSKDVSEAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420
QY 421 ISFOGLCMLLDFVSMVYVYHFRKASIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 480
Db 421 ISFOGLCMLLDFVSMVYVYHFRKASIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 480
QY 481 CILLRWARLLGFATVYGTVTLKLRVLFVLSRTAQRIPIYMTGGVRMRMLAVILLVWF 540
Db 481 CILLRWARLLGFATVYGTVTLKLRVLFVLSRTAQRIPIYMTGGVRMRMLAVILLVWF 540
QY 541 LGWTSSVCQNLEKQISLIGQKTSKGLGRSIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 600
Db 541 LGWTSSVCQNLEKQISLIGQKTSKGLGRSIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 600
QY 601 RTVPSAFHPRYMAYVHNELIISAFIHTIRVFLASRLQSDMMLMLYFAHTHLTVITIG 660
Db 601 RTVPSAFHPRYMAYVHNELIISAFIHTIRVFLASRLQSDMMLMLYFAHTHLTVITIG 660
QY 661 LLLIPKFSHSSNNPRDDIATEAYEDELDMGRSGSYLNSINSAWSEHSLDPEIDELK 720
Db 661 LLLIPKFSHSSNNPRDDIATEAYEDELDMGRSGSYLNSINSAWSEHSLDPEIDELK 720
QY 721 LYAQLEIYKRRKMITNPHLOKRCSSKGLGRSIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 780
Db 721 LYAQLEIYKRRKMITNPHLOKRCSSKGLGRSIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 780
QY 781 TAKGTALIRKNPESNGTKSKEETLNKRVFSLKSHSTYDHRDQTEESSSLPTESEQ 840
Db 781 TAKGTALIRKNPESNGTKSKEETLNKRVFSLKSHSTYDHRDQTEESSSLPTESEQ 840
QY 841 EETENSTLESLSGKKLTQKLKEDSEAESESTESVPLVCKSASAHNLSSEKKTGHPRTSMQ 900
Db 841 EETENSTLESLSGKKLTQKLKEDSEAESESTESVPLVCKSASAHNLSSEKKTGHPRTSMQ 900
QY 901 KLSVTASAKETLGLAGKTQTAGVEERTKSKPLPKOKETNRNHSNDSNTETKDPAPON 960
Db 901 KLSVTASAKETLGLAGKTQTAGVEERTKSKPLPKOKETNRNHSNDSNTETKDPAPON 960
QY 961 SNPAEPRKPKSGIMKQORVNPPTANSDLNPGTTQMKDNFDIGEVCPWEVYDLTPGPVP 1020
Db 961 SNPAEPRKPKSGIMKQORVNPPTANSDLNPGTTQMKDNFDIGEVCPWEVYDLTPGPVP 1020
QY 1021 SESKVKQHSIVASEMEKNPTFSLKESHKPKAAEVCOQSNOKRIDKAEVCLWESQGS 1080
Db 1021 SESKVKQHSIVASEMEKNPTFSLKESHKPKAAEVCOQSNOKRIDKAEVCLWESQGS 1080
QY 1081 ILEDEKLLISKTPVLPRAKEENGOPRAANVACQSGEELPPKAVASKTENENLQIGHQ 1140
Db 1081 ILEDEKLLISKTPVLPRAKEENGOPRAANVACQSGEELPPKAVASKTENENLQIGHQ 1140
QY 1141 EKKTSSEENVRGYSNNSNFQOPLTSRAEVCWPEFETPAQPNAGRSVALPASSALSANK 1200
Db 1141 EKKTSSEENVRGYSNNSNFQOPLTSRAEVCWPEFETPAQPNAGRSVALPASSALSANK 1200
QY 1201 IAGPRKEEIWDSFKV 1215
Db 1201 IAGPRKEEIWDSFKV 1215
```

RESULT 2  
US-09-775-181-4  
; Sequence 4, Application US/09775181  
; Patent No. US20020038013A1

GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; TITLE OF INVENTION: NO. US20020038013A1el Human Membrane Proteins and  
; FILE REFERENCE: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0129-USA  
; CURRENT APPLICATION NUMBER: US/09/775,181  
; CURRENT FILING DATE: 2001-02-01  
; PRIOR APPLICATION NUMBER: US 60/180,414  
; FILING DATE: 2000-02-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 599  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-775-181-4

Query Match 48.9%; Score 3119.5; DB 10; Length 599;  
Best Local Similarity 98.5%; Pred. No. 1.2e-196;  
Matches 585; Conservative 2; Mismatches 2; Indels 5; Gaps 1;

```
QY 1 MGAMAYPLLICLLLAQLGLGAVGASRDPOGRPDSPRTPKGPQAQOPGRASASDSSAP 60
Db 1 MGAMAYPLLICLLLAQLGLGAVGASRDPOGRPDSPRTPKGPQAQOPGRASASDSSAP 60
QY 61 WSRSTDGTLIAQLAEVPMVASYLYTGDHQLKCRANSGRYELAGLPGKWPALASAP 120
Db 61 WSRSTDGTLIAQLAEVPMVASYLYTGDHQLKCRANSGRYELAGLPGKWPALASAP 120
QY 121 SLHRALDTLTHATNPLNVMQSKREONLQDDLDWYQALVWSLLEGEPSISRAITEST 180
Db 121 SLHRALDTLTHATNPLNVMQSKREONLQDDLDWYQALVWSLLEGEPSISRAITEST 180
QY 181 DLSLAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPRLHRRGN 240
Db 181 DLSLAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPRLHRRGN 240
QY 241 QGPRGLGHSWRKDGKSHFKWSPPYLECENGSYKPGWLVTLSAIYGLQPNLVPEF 300
Db 241 QGPRGLGHSWRKDGKSHFKWSPPYLECENGSYKPGWLVTLSAIYGLQPNLVPEF 300
QY 301 RGVMKVDINLQKVDIDQCSSDGFSGTHKCHLNNSCMPKIGLGFVLGAYECICAGFYH 360
Db 301 RGVMKVDINLQKVDIDQCSSDGFSGTHKCHLNNSCMPKIGLGFVLGAYECICAGFYH 360
QY 361 PGVLPVNNFRRRGPDQHIISGSKDVSEAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420
Db 361 PGVLPVNNFRRRGPDQHIISGSKDVSEAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420
QY 421 ISFOGLCMLLDFVSMVYVYHFRKASIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 480
Db 421 ISFOGLCMLLDFVSMVYVYHFRKASIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 480
QY 481 CILLRWARLLGFATVYGTVTLKLRVLFVLSRTAQRIPIYMTGGVRMRMLAVILLVWF 540
Db 481 CILLRWARLLGFATVYGTVTLKLRVLFVLSRTAQRIPIYMTGGVRMRMLAVILLVWF 540
QY 541 LGWTSSVCQNLEKQISLIGQKTSKGLGRSIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 594
Db 541 LGWTSSVCQNLEKQISLIGQKTSKGLGRSIRASGLILLETILFGSLLLYFPVVIYFEPSTFR 594
```

RESULT 3  
US-10-027-923-5  
; Sequence 5, Application US/10027923  
; Patent No. US20020142330A1  
GENERAL INFORMATION:  
; APPLICANT: Briann Gaither Bates  
; APPLICANT: Kamalaka Gulukota  
; APPLICANT: Yuhong Xie  
; APPLICANT: Janet Elizabeth Paulsen  
; TITLE OF INVENTION: NOVEL GLUTAMATE RECEPTOR MODULATORY PROTEINS AND

	Matches	207; Conservative	135; Mismatches	383; Indels	387; Gaps
Qy	219	ETEFWHLGRKWRPHLHRRGPNQGPRGLGHWSR-----RKDGLGDGDKSHFKNSPPYLECE	273	: :	:
Db	333	DVKWFDDYYLKLRPETNHRNP-----WFQEWMHRFOCRLEGFPQENSKYNKT-----CN	382	: :	:
Qy	274	NG-----SYKPGWLVTLSAIYIGL---QPNIIVPEFRV-----MKVDINLQ	311	: :	:
Db	383	SSLTAKTHHVQDSKMGEFVNATYSMAYGHLHMOMSLCPGYAGLCDAMKPIDGRKLLESIM	442	: :	:
Qy	312	KVIDIDCSSDGWFSGTHKHCHLNNSCMP1KGIGLVFLGAYECICAKGFHPGVLPVPNNFR	371	: :	:
Db	443	KTNTGTVGSGDTILFDENGDSGRYEIMNPKMG-----KDYFYINVGSMDNGEL	492	: :	:
Qy	372	RGPDOHT-----SGSTRKDVEE-----AYVCLPCREG-----CPFC	402	: :	:
Db	493	KMDDDEWSSKSNIIRSVCSEPCKEGOIKVIRKEGVSCCWTCTPCKENEYVFDEYTKAC	552	: :	:
Qy	403	-----ADD--SPCFVGEDKYLR-----LAITSFGOLCMLLDFVSMLVVYHFRAKSIRA	449	: :	:
Db	553	QLGSWPTDDLTTGCDLIPVQYLRGWDPEPIAAVVFACLGILLATFTVTVFIYRTDPVVK	612	: :	:
Qy	450	SGLILETILFGSLLLYFPVVIILYPBPSTFRCELLRWARLLGFATVYGTVLKLRVLKV	509	: :	:
Db	613	SSRELCVIIIAGLCGLYCTFCCLIAKPQICYLQRIGTIGLSPAMSYSALVTKTNRIARI	672	: :	:
Qy	510	-----FLSRTAORIPYMTGGVRMRMLAVILLVVFWFLIGWTSVCONLEKQ	555	: :	:
Db	673	LASGKKIKCTKKRPMSACAQ-----LVTAFTILI-----CIQUGIIL	708	: :	:
Qy	556	ISL-----IGOGKTSDDLIFNMCLIDRDYMTAVA-EFLFLLMGVLYCAVRTVPSAF	607	: :	:
Db	709	VAFIMEPPDIIMHDYSPISREVLICNTNLGVVTPLYNGLLIILSCTFYAPKTRNVANPF	768	: :	:
Qy	608	HEPRYMAVAVHNELIISAIFHTIRFVLASRLQSDWMLMLYFAHTHTVTVTIGLLIPK-	666	: :	:
Db	769	NFAKYTAFTWYTTCITWLAEVTFVFE-----CSNYKITMCESVISATVALGWMEVDKV	822	: :	:

Qy	667	---F\$HSSNNPRDIIATE-AYEDELMDGRSGYL--NS\$INSANSEHSLOPEDIRDELKK	720
	:	:	:
Db	823	YIILAKPERNVRSAFTTSTVVRMHVGDGKSSAA\$KSS\$SLNLMKRRGSSGETLR----	877
	:	:	:
Qy	721	LYAQLETYK\$K\$K\$K\$M\$N\$PHLQK\$K\$K\$GLGR\$IMRRITETPEV\$YQ\$CKEDK\$GADHG	780
	:	:	:
Db	878	-----YKORRLAQ\$K\$E\$T-----ECF-----	893
	:	:	:
Qy	781	TAKGTALIRNPP\$ESSGN\$TK\$KEETLKNRVP\$LK\$K\$H\$TYDHYRDQTE\$SS\$LPTE\$SQE	840
	:	:	:
Db	894	TPKG-----SMNGNGRA-----TWSSNGK\$SVTWAQN	920
	:	:	:
Qy	841	EETEN\$T\$LES\$G\$K\$LTOKL-----KED\$E-----AESTESVPLCK\$SA\$H\$N\$LS	886
	:	:	:
Db	921	EK\$SR-----GQHLW\$OR\$LI\$H\$K\$KN\$PNQ\$TAVI\$P\$P\$K\$TES\$RGL\$GAG\$GAG\$G\$	972
	:	:	:
Qy	887	SEKKTG-----H\$P\$T\$M\$LQ\$K\$S\$VIA\$AK\$E\$K\$T\$LG\$L\$A\$G\$K\$T\$Q\$T\$A\$G\$V\$E\$R\$T\$K\$S\$Q	933
	:	:	:
Db	973	GVGATG\$GAG\$CAG\$GPG\$P\$DAG--PRALYDVAE\$E\$H\$P\$A\$P-----R\$P\$S\$P\$S	1020
	:	:	:
Qy	934	PLPKDKETN\$N\$H\$N\$S\$N\$T\$K\$D\$P\$A\$Q--N\$N\$P\$A\$E\$P\$R\$K\$Q\$K\$G\$IM\$K\$Q--Q\$R\$V\$N\$P\$T\$A\$N--S\$D	989
	:	:	:
Db	1021	PI-----STL\$H\$R\$A\$G\$A\$S\$R\$T\$D\$D\$V\$P\$S\$H\$E\$P\$A\$R\$S\$S\$Q\$G\$S--L\$M\$E\$Q\$T\$S\$V\$V\$T\$R\$T\$A\$N\$I\$S	1074
	:	:	:
Qy	990	LN\$P\$C\$T\$Q\$M\$D\$N\$F\$D\$IG\$E\$V\$C\$P\$W\$E\$Y\$D\$LT\$P\$G--P\$V\$P\$S\$E\$K\$V\$Q\$K\$H\$V\$S\$IV\$A\$E\$K\$N\$P\$T\$F\$S\$L\$K\$E	1046
	:	:	:
Db	1075	LN-----SMML\$T\$A\$P\$-----S\$P\$G\$A\$P\$LC\$S\$Y\$LI\$P\$K\$E\$I\$Q\$L-----P\$T-----	1107
	:	:	:
Qy	1047	K\$H\$H\$K\$P\$K\$A\$E\$V\$C\$Q\$S\$N\$O\$K\$R\$ID\$K\$A\$E\$V\$C\$L\$W\$E\$S\$Q\$G\$S\$T\$E\$K\$E\$K\$L\$K\$T\$P\$V\$L\$P\$E\$R\$A\$K\$E\$N\$G\$G--	1105
	:	:	:
Db	1108	-----TW\$T\$F\$A\$E\$I\$Q\$L\$P\$A\$E\$V\$T\$G\$A	1127
	:	:	:
Qy	1106	Q\$P\$R\$A\$A\$N\$Y\$C\$A\$G\$-----Q\$S\$E\$E\$L\$P\$P\$K\$A\$V\$A\$K\$T\$E\$N\$E\$N\$L\$N\$Q\$I\$H\$Q\$E\$K\$T\$S\$S\$E\$E\$N\$V\$R\$G\$Y\$S\$N\$N\$N\$F	1161
	:	:	:
Db	1128	O\$P\$A\$Q\$A\$G\$A\$D\$A\$E\$S\$P\$A\$G\$P\$E\$A\$A\$A\$K\$P\$D\$E\$L\$V\$A\$L-----T\$P\$P\$S\$P\$R\$D\$S\$V\$D\$S\$G\$S\$T--	1178
	:	:	:
Qy	1162	Q\$Q\$P\$L\$T\$S\$R\$A\$E\$V\$C\$P\$W\$E\$F\$T\$T\$A\$Q\$P\$N\$A\$G\$R\$S\$V\$A\$L\$P\$A\$S	1193
	:	:	:
Db	1179	-----T\$P\$N\$S\$P\$Y\$S\$E\$A\$L\$C\$I\$P\$S\$	1194

## RESULT 5

US-09-864-761-35612  
; Sequence 35612, Application US/09864761  
; Patent No. US20020048763A1

GENERAL INFORMATION:

: APPLICANT: Penn, Sharron G.  
 : APPLICANT: Rank, David R.  
 : APPLICANT: Hanzel, David K.  
 : APPLICANT: Chen, Wensheng  
 : TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 : TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

```

FILE REFERENCE: Aecomica-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24363.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/006656
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006659

```

```

> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00665
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00668
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00663
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00662
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00661
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: PCT/US01/00670
> PRIOR FILING DATE: 2001-01-30
> PRIOR APPLICATION NUMBER: US 60/234,687
> PRIOR FILING DATE: 2000-09-21
> PRIOR APPLICATION NUMBER: US 09/608,408
> PRIOR FILING DATE: 2000-06-30
> PRIOR APPLICATION NUMBER: US 09/774,203
> PRIOR FILING DATE: 2001-01-29
> NUMBER OF SEQ ID NOS: 49117
> SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
> SEQ ID NO 35612
> LENGTH: 1325
> TYPE: PRT
> ORGANISM: Homo sapiens
> FEATURE:
> OTHER INFORMATION: MAP TO AP000511.1
> OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
> OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
> OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
> OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
> OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
> OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
> OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
> OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
> OTHER INFORMATION: EST_HUMAN HIT: AW867076.1, EVALU0 1.00e-103
> OTHER INFORMATION: EST_HUMAN HIT: AW867076.1, EVALU0 1.00e-103
> OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALU0 5.00e-63
> US-09-864-761-35612

```

Query Match 3.0%; Score 188.5; DB 10; Length 1325;  
Best Local Similarity 22.1%; Pred. No. 0.00035;

Qy	669	HSSNPRDDIAT	EAYEDELDMG	SGSYLNSS	INSAWSEHSLDP	EDIRDELKLYAQLEIY	728							
		:	:	:	:	:								
Db	80	HKSSTDNHEAP	PTSEENSNOG	DPMRN	-----QRSVDPAD	STTTTHKESAG	126							
Qy	729	KRKXMIINPH	LQKRCSSKGL	GRSMRRIT	ETPETSQCSK	EDKEDGADHTAKGTALI	788							
Db	127	--KKHITPAP	K-SKINCRK	STTKGS	-----TVTR--K	SDYTG--RPLEKSMSTL	168							
Qy	789	RKNPPES	-----SGNT-	GSKBETLKN	RVSFLKKSH	STVHDVDRDESSSLPTES	838							
Db	169	DKTSSSHKTT	TFHNSGNSQ	TQKSTSP	EKITAASKT	-----TYKTTGTPEBS	218							
Qy	839	QEEB	-----TTENSTL	ESLGGK	UKLOKLED	SEASTESVPLVCKSASAH	884							
		:	:	:	:	:								
Db	219	EKTEDSR	TTVASDKLL	TKTKN	-IQETISANEL	TQSLAEPTHEGG	-----RTANENN	269						
Qy	885	LSS	-EKKTGHP	TSMLQK	SLVSIASAK	EKTLGLACK	TQTAGVEERTKSOK	--PLPKDKE	940					
Db	270	TPSPAEPTE	NRERT	-----ANENT	TLSPAEP	TEN	-----REKTANENT	APPAGPT	315					
Qy	941	TNRHNSD	NETDKPA	-----PQNSN	PABEP	RKPKQSG	IMKQOORVNP	TTANSDLNP	992					
Db	316	ENREMTAN	ENT-LT	FPAEPTE	HGERTAN	ENTTSPA	EPTEHG	--ERTANENT	TTPSPAEP	371				
Qy	993	-----G	TTQMKDN	FDIG	VECPWEV	YDLTPGP	-----VPS	-----	102					
		:	:	:	:	:	:	:						
Db	372	TEHGERT	PFA	NDKTTSS	AEASTEH	GERTPLAN	ENTTSPA	EPTENR	ERTANENT	TTPSPAG	431			
Qy	1022	-----PSK	VOKH	SVI	VA	SEK	NPTESL	KPKSHHK	PKA	AVCOOS	NOKR	TRDKAEVCI	WPS	1074





```
Db 208 EEVS-----REIVSTSTAPSRIVER-----GKTKQVIKEQPETGVHKKDVS 252
QY 870 TESVPLVCKASAHNLSBKKTGHP-----RTSMLQKSLSVIASAKEKTL 914
Db 253 GAIVEPAIOPELPEAVVSDK--GEPEVQTLPEAVVYDKGETEVQPESDTVVSDKGEPE 310
QY 915 GLAGKTQTAGVEERTKSKQPLPKDKETNRHNSNDTETKDPAPQNSN-----PA 964
Db 311 QVAPLPEYKGNIEQVQPEPVKTEKQGPKEETEVPVKTEETPVNPNEGTGTGSIQEA 370
QY 965 EEPKPKQSGIMKQORVNTTANSNDLPCTTQMKONFDIGEVCPWEVYDLTGPVPVSESK 1024
Db 371 ENVPQPAESTTNSKVSPTSKN-----TGEVSSN-----PSDS- 406
QY 1025 VOKHVSIVASEMEKNPTFSLKPKSHHKPKAAAEVCCQOSNKRDKARVCLWES--QGOSIL 1082
Db 407 -----TTSVGES--NKPEHNSKNSEKTEVEVPVNPNEGTVGTSNQ 448
QY 1083 EDEKLLIKSTPVLPERAKEENGQPPRAANVCAGQ-----SEELPP-----KAVASK 1128
Db 449 ETEK-----PVQPAKETQNSG--KIANENTGEVSNKPSDKPPVEESNQPEKNGTATK 500
1129 TEN-ENLNOIGHQEKKTS--SSEENVRGSYNSNN-NFQOPLTSRAEVCP 1173
Db 501 PENSNGTTSNGQTEPSPNGSNTEDVSTESNTSNGNEEIKQENELDP 550

RESULT 9
US-09-864-761-36182
; Sequence 36182, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aomicla-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
```

```
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36182
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005529.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
; OTHER INFORMATION: SWISSPROT HIT: P12036, EVALUE 2.00e-33
; OTHER INFORMATION: EST_HUMAN HIT: BF445754.1, EVALUE 3.00e-22
US-09-864-761-36182

Query Match 2.7%; Score 174; DB 10; Length 617;
Best Local Similarity 20.6%; Pred. No. 0.0011;
Matches 101; Conservative 79; Mismatches 229; Indels 82; Gaps 17;

QY 754 IMRRITEIPETVSROCSKEDKGAHGHTAKGTALIRKNPP-ESSGNTGKSKEETLKNRVF 812
Db 63 VTEEVTEEEKEAKEEGKEEGEEAEGETTKSPABEAAKSPAEKASPVKEEK 122
QY 813 SLKKSNS-----TYDHRDQTESSSLPTESQEEETTE-NSTLESLSGKILT 858
Db 123 SPAEAKSPEKEEAKSPAEVKSPEKAKSPAEKASPEKEEAKSPAEVKSPEKAK 182
QY 859 QKLKEDSEAEESTESVPLVCKSASAHNLSSEKKTGHPRTSMLQKSLSVIA--SAKEKTLGL 916
Db 183 SPAKEEAKSPAEKASPEKAKSPVKEEAKSPAEKASPAEVKSPEKAKSPTKEE 242
QY 917 AGKTOTAGVEER-----TKSOKPLPKDKETNRHNSNDTETKDPAPQNSNPAEPRKP 970
Db 243 AKSPEKASPEKEEAKSPEKAKSPVKAEEKSPVKAEEKSPVKAEEKSPVKEEKSP 302
QY 971 QKS-GIMKQORVNTTANSNDLPNGTQMKDNFDIGEVCPWEVYDLTGPVPVSESKVQKHV 1029
Db 303 EKAKSPVKEEAKSPEKAKS-----PVKEEAKTPEKA 333
QY 1030 SIVASEMEKNPTFSLKESHHPKAAAEVCCQOSNQRIDKAEVCLWESQCSQSILEDEKLLI 1089
Db 334 KSPVKEEAKSPE-----KAKSPEKAKTLDV--KSPEAKTPAKE-----EARSADKFPPEK--- 381
QY 1090 SKTPVLPERAKEENGQPPRAANVCAGOSELPKPAVASK--TENENLNOIGHQEKKTSS 1147
Db 382 AKSPVKEEVKSPEKAKSPLKADAKAPE-KEIPKKEEVKSVPVKEEKQPQVKEKPPKAE 440
QY 1148 EENVRGSYNSNNFQOPLTSRAEVCPWEPEPTP---AQPNAGRSVALPASSALSANKIAP 1204
Db 441 EEKAPATPKT---EKKDSKKEAP-KKEAPKPVKEEKKEPAVEKPKESKVEA----- 489
QY 1205 RKEEIWDSFKV 1215
Db 490 KKEAEADKKV 500

RESULT 10
US-09-839-996-6
; Sequence 6, Application US/09839996
; Publication No. US2003000910A1
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
```

NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/839,996  
FILING DATE: 20-Apr-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/296,791  
FILING DATE: 25-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Trecartin, Richard F.  
REGISTRATION NUMBER: 31,801  
REFERENCE/DOCKET NUMBER: A-59941/RET/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1848 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-839-996-6  
Query Match 2.6%; Score 168.5; DB 9; Length 1848;  
Best Local Similarity 18.4%; Pred. No. 0.012;  
Matches 128; Conservative 111; Mismatches 266; Indels 189; Gaps 28;  
QY 626 IFHTIRFVLSKRL---QSDW-----MLMLYFAHFLTV-----TVTIGLLLLPK 666  
DB 846 LFTQISGTSQVNLKSHHHLTGNSVNVNLNGLHILNAQNDANKVTTYNTLTVNS 905  
QY 667 FSHS-----SNNPRDIAT-----EAYEDLMGRSGSYLNSI 700  
QY 906 LSGSGFYVWDFTNKSKVYVNVKSGATGNFTLQVADKTGEPNHNELTLFDASNATRN 965  
QY 701 NSAWSHSLDPEDIRDELKLYAQLEIY-----KRKKM-----ITNPHLOKRCCKGL 750  
DB 966 EYTLANGSVDRGAWKYKLNVRGVDLYNPEYKRNQIVDTNITPNDIQADAPSAQSN 1025  
QY 751 GRSIMRRITEIP---ETVSRQSKEDKAGDHGTAKGTALIRKNPPESGNTGSKETL 807  
DB 1026 NEEIARVETPVPPAPAPATESAIASEQEPTRPETAQ-----PAMEETNTANSTETAP 1077  
QY 808 KNRVSLKSHSHYDHRDQTE--ESSSLPTESQEE-----ETTENSTLESGLKLTOKL 861  
DB 1078 KS-----DTATQTNPNSESVSETTEKVAENPPQNETV-AKNEQATEPT 1123  
QY 862 KEDSAESTESVPLVCKSASAHNLSSEKKTGHPTSMLOQ-----KSLSVIASAKEKTLG-- 915  
DB 1124 PONGEVAKEDQPTVEANTQTNATQSEKGTETQTAETKSEPTESVTSYSENQPEKTVSQS 1183  
QY 916 -----LAKTQTAGVE-----ERTKSKPLPKDKETNRNNSNSDNTETKDP- 956  
DB 1184 TEDKVVEKEEAKVETETQKAPQVTSKEPPKQAPAPAEVPTDNTNAEEAQAQLOQTPT 1243  
QY 957 -----APQNSNPAEERPKQKSGIMQOORVNPNTANSDLNPGTQMKDNFDICEVCPW 1009  
DB 1244 TVAAAEITSPNKPAEETQQPS-----BKTNAPVY----- 1274  
QY 1010 EYDITPGVPFVPSKQKHSIVASEMEKNPTFLSLKEKSHHKPAAEVCCQSQNOKRIDKA 1069

DB 1275 -----PVVSENTATQ-----PTETEETAKVE-KEXTQEVP-----QVASQESPKQEQPA 1317  
QY 1070 EYCLWESQGSSTLEDEKLLISRTPLPE---RAKENGQOPRAANYCAGQSEELPPKAVA 1126  
DB 1318 AKPQAQTKPQAPAPARENVLTTKNVGEPPQQAQPTQSTAVPTTGTETAANSKPAKPOQA 1377  
QY 1127 S-KTE--NENLAQIGHOE--KKTSSSEENVRSYSSNNFQOP-----LTSR 1168  
DB 1378 KPQTEPARENVTNTKBPQSQTSATVSTEQPAKETSSNVQEPAPENSIINTGSATMTTET 1437  
QY 1169 AEWCPWEFTPAQPNAGRVALPASSALSANKIA 1202  
DB 1438 AE----KSDKPQMEIVTENDROPEANTVADNSVA 1467  
RESULT 11  
US-09-902-432-4  
; Sequence 4, Application US/09902432  
; Patent No. US20020160002A1  
; GENERAL INFORMATION:  
; APPLICANT: Irwin H. Gelman  
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE  
; FILE REFERENCE: A30558-A-FWC-A 070156 0597  
; CURRENT APPLICATION NUMBER: US/09/902,432  
; CURRENT FILING DATE: 2002-04-08  
; PRIOR APPLICATION NUMBER: 08/978,277  
; PRIOR FILING DATE: 1997-11-25  
; PRIOR APPLICATION NUMBER: 08/665,401  
; PRIOR FILING DATE: 1996-06-18  
; PRIOR APPLICATION NUMBER: 08/635,121  
; PRIOR FILING DATE: 1996-04-19  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 1596  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-09-902-432-4

Query Match 2.6%; Score 164; DB 9; Length 1596;  
Best Local Similarity 22.1%; Pred. No. 0.019;  
Matches 135; Conservative 101; Mismatches 257; Indels 118; Gaps 34;  
QY 664 IPKFSSHNNPRDDIATAYE--DELDMGSRGSGYLNSSINSAMSEHSLLDPEDIRDELKLY 722  
DB 239 IPLQAESDQAAEEAAKDEGEKQKEPTKSPESPSPVNS-----ETTSSFKKFF 288  
QY 723 A-----QLEIYKRKKMITNPHLOKRCCKKGLGRSIMRRITEIPETVSRQSKEDKE 775  
DB 289 THGNAGWRKKTSFKKSK--EDDLTAERKQEA-----EKVDEEKEKTEPASEEQEP 340  
QY 776 GADHGTAGTALIRK--NPPES--GNTCKSKEE---TLKNRVFSLK-KSHS---TYDHVR 825  
DB 341 AEDTQOARLSADYKVELPLEDQVDLEASSEEKCAPLATEVFEDEKMAHAEVVAEVS 400  
QY 826 --DOTFESSSLPTESQEEETTENTSTLESGLKLTOKLKEDESAESTESVPL---VCKSA 880  
DB 401 TVEKTEEEGGGGAEGGVVVE--GTGESLPPKLAEPQEVPOEAEPABELMKSRMCVSG 459  
QY 881 SAH---NLSSEKT--CHP-----RTSML--OKSLSVIASAKETLCLAGKTQTAGVEE 927  
DB 460 GDHQLTLDLSPPEETLPHKPHGIVSEVEMLSQBRIRKVGQSPKLKLFSSSGKLKLSG--K 517  
QY 928 RTKSKPLPKDKETNR--NHSNSDNTETKDPAPONSNAEPRKPKQKSGIMQOORVNPPTA 986  
DB 518 KQKGRGGGGDEEPEVQHIHTESPEADEQKGESS--ASSPEEPEETCLEK---GPLEA 573  
QY 987 NSD--LNPQTOMKDNFDIGEVCPWEVYD--LTPGPVPSKQKHSIVASEMEKNPTF 1042  
DB 574 PDGGEAEGETTSDGEKKREG-ITPWASFKMWTP-----KKRV-----RRPSE 615







Db 873 IHVKTSDDDHSGKIKSEI---EPKELTEESPLTD-----KKTESAAIGDKNHESV 920  
QY 1053 KAAEVQ-----QSNQKRIDKAEVCLWESQ-----QSILEDKLLISKTPVLPERAKEE 1102  
Db 921 KSADIFQSEIHNSDNDRI-VSESVQDSSGSSMSTESIRTNDKDKTSEDIAPI-----SI 975  
QY 1103 NGGQPRAAWVAGQSEELPPKAVASK-TENENLNQIGHQEKKTSSSEENVRGSYNSSNNF 1161  
Db 976 NGHEKIGSSADDRGSED---KSIIDKDSFENFNKSSHSDIKQSDNEGST--DYESLTE- 1029  
QY 1162 QOPLTSRAEVCPEFETPAQPNAGRSVALPASSALSANKIAGRKEEI 1209  
Db 1030 ESPKGDLESVPSSIDMDLKPNI-----KSPVTSFDHVDSPNISEL 1070

Search completed: February 13, 2003, 02:31:08  
Job time : 36.4146 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: February 15, 2003, 04:23:28 ; Search time 30.7095 Seconds  
(without alignments)  
2599.103 Million cell updates/sec

Title: US-09-775-181-4  
Perfect score: 3199  
Sequence: 1 MGAMAYPLLLCLLQAQLGLG.....YMTAGMWSLVSYDGLTIFQ 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_101002:\*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	3199	100.0 599 22	AAE06643 Human G-protein co
2	3119.5	97.5 1215 22	AAE06642 Human G-protein co
3	339	10.6 669 22	ABB70858 Drosophila melanog
4	339	10.6 669 22	ABB70858 Drosophila G-prote
5	238.5	7.5 176 22	ABB67185 Drosophila melanog
6	161.5	5.0 976 22	ABB65498 Drosophila melanog
7	149	4.7 1677 22	ABB65976 Drosophila melanog
8	141.5	4.4 974 21	AA145000 Human calcium sens
9	139.5	4.4 917 20	AA149126 ChimERIC hCAR/hmG1
10	139.5	4.4 917 23	AAO15092 ChimERIC human hCA

11	139.5	4.4	1276	20	AA149127	phCar/hmGluR2*Gqi5
12	139.5	4.4	1276	23	AAO15093	Human phCar/hmGluR
13	136	4.3	901	21	AA145001	Human calcium sens
14	134.5	4.2	1305	22	AA186161	D. melanogaster GA
15	133.5	4.2	1026	18	AA132059	Dogfish shark kidn
16	133.5	4.2	1079	19	AA154847	Rat kidney calcium
17	133.5	4.2	1079	19	AA138275	Rat kidney cell ca
18	133.5	4.2	1079	20	AA141781	Rat parathyroid ca
19	133.5	4.2	1079	20	AA14928	Rat kidney extrace
20	133.5	4.2	1079	20	AA189566	Rat parathyroid ca
21	133.5	4.2	1079	21	AA151828	Rat calcium recept
22	133.5	4.2	1079	23	AA147823	RakCar3A. rattus
23	132	4.1	1001	21	AA144999	Human calcium sens
24	132	4.1	1027	23	ABB78761	Dogfish shark kidn
25	132	4.1	1027	23	AA176004	Shark kidney calci
26	130	4.1	1305	22	ABB59497	Drosophila melanog
27	130	4.1	1305	22	AA138927	Drosophila G-prote
28	127.5	4.0	840	21	AA145021	Rat sensory transd
29	127.5	4.0	840	21	AA145026	Rat sensory transd
30	127.5	4.0	840	21	AA145027	Rat sensory transd
31	127.5	4.0	840	21	AA145028	Rat sensory transd
32	127.5	4.0	1138	23	AAO21502	1138-mer rat prote
33	127	4.0	1078	17	AA111889	Parathyroid calciu
34	127	4.0	1078	19	AA154846	Human parathyroid
35	127	4.0	1078	19	AA138274	Human parathyroid
36	127	4.0	1078	20	AA128840	Human calcium rece
37	127	4.0	1078	20	AA141780	Human parathyroid
38	127	4.0	1078	20	AA189565	Human parathyroid
39	127	4.0	1078	21	AA151827	Human calcium rece
40	127	4.0	1078	21	AA170325	Human wild type ca
41	127	4.0	1078	22	AA174391	Protein encoded by
42	127	4.0	1078	23	AA147822	HuCar4.0. Homo sa
43	124	3.9	1078	22	AAO02195	Cynomolgous monkey
44	123	3.8	1085	19	AA154844	Bovine parathyroid
45	123	3.8	1085	19	AA138272	Bovine parathyroid

ALIGNMENTS

RESULT 1  
AAE06643  
ID AAE06643 standard; Protein; 599 AA.  
AC AAE06643;  
XX  
XX  
DT 16-OCT-2001 (first entry)  
XX  
XX  
DE Human G-protein coupled receptor (NGPCR) #2.  
KW Human: G-protein coupled receptor; NGPCR; gene therapy; cancer; diabetes;  
KW inflammation; immune disorder; heart disease; obesity; coronary disease;  
KW metabolic disorder; physiological disorder; therapeutic; drug screening;  
KW signal transduction; behavioural disorder.  
XX  
XX  
OS Homo sapiens.  
XX  
PN WO200157086-A2.  
XX  
XX  
PD 09-AUG-2001.  
XX  
XX  
PF 01-FEB-2001; 2001WO-US03648.  
XX  
XX  
PR 04-FEB-2000; 2000US-0180414.  
XX  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
XX  
PI Donoho G, Hilbun E;  
XX  
XX  
DR WPI: 2001-488870/53.  
DR N-PSDB; AAD12293.  
XX  
XX  
PT Novel isolated polynucleotides encoding human G protein coupled

PT receptor (NGPCR), useful as probe and for treating disease involving  
PT GPCR, detecting mutant NGPCR or abnormally expressed NGPCRs for  
PT diagnosing disease  
XX  
XX  
XX Claim 4; Page 70-71; 73pp; English.  
XX  
XX The present sequence is human G-protein coupled receptor (NGPCR) which  
CC is a 7 transmembrane protein of G-protein coupled receptor (GPCR) super  
CC family. NGPCR is expressed in hypothalamus, foetal brain, brain,  
CC cerebellum and testis. NGPCR spans the cellular membrane and is involved  
CC in signal transduction pathways. Human NGPCR is useful for treating  
CC obesity, inflammation, immune disorders, diabetes, heart and coronary  
CC disease, metabolic disorders and cancer. NGPCR's are also used as  
CC reagents in assays for screening compounds that are useful in the  
CC therapeutic treatment of physiological and behavioural disorders. NGPCR  
CC is useful for identifying compounds that modulate NGPCR gene expression  
CC and/or gene product activity. NGPCR DNA is also useful for the diagnostic  
CC and prognostic evaluation of disorders related to NGPCR function.  
CC Nucleotide constructs encoding functional NGPCR and mutant NGPCR variants  
CC are used in gene therapy. NGPCR probes are also useful for the  
CC identification, selection and validation of novel molecular targets for  
CC drug discovery.  
XX  
XX  
XX Sequence 599 AA;

Query Match 100.0%; Score 3199; DB 22; Length 599;  
Best Local Similarity 100.0%; Pred. No. 9.1e-300; Indels 0; Gaps 0;  
Matches 599; Conservative 0; Mismatches 0;

QY 1 MGAMAYPLLCLLLAQLGLGAVGASRDQGRDPSRPTPKGPKHAQPGGRASDSSAP 60  
DB 1 MGAMAYPLLCLLLAQLGLGAVGASRDQGRDPSRPTPKGPKHAQPGGRASDSSAP 60  
QY 61 WSRSTGDTTLAOKLAEEVPMVASYLTGDSHQLKRANCSGRYELAGLPKWPALASAH 120  
DB 61 WSRSTGDTTLAOKLAEEVPMVASYLTGDSHQLKRANCSGRYELAGLPKWPALASAH 120  
QY 121 SLHRALDTLTATNLFNVLNMQSNKREONLQDDLDWYQALVMSLLEGEPSISRAAITFST 180  
DB 121 SLHRALDTLTATNLFNVLNMQSNKREONLQDDLDWYQALVMSLLEGEPSISRAAITFST 180  
QY 181 DLSAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPPLHRRGN 240  
DB 181 DLSAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPPLHRRGN 240  
QY 241 QGPRGLGHSWRKDLGGDKSHFKWSPPLYCEGNSYKPGWLVTLSSAIYGLQNLVPEF 300  
DB 241 QGPRGLGHSWRKDLGGDKSHFKWSPPLYCEGNSYKPGWLVTLSSAIYGLQNLVPEF 300  
QY 301 RGVKMDINLQKVDIDQCSDDGWFSGTHKHLNNSECMPIKGLGFLVGLAYECICRAGFYH 360  
DB 301 RGVKMDINLQKVDIDQCSDDGWFSGTHKHLNNSECMPIKGLGFLVGLAYECICRAGFYH 360  
QY 361 PGVLPVNNRRRGPDPQHISGSKDYSEAYVCLPREGCPFCADDSPCFVQEDKYLRLAI 420  
DB 361 PGVLPVNNRRRGPDPQHISGSKDYSEAYVCLPREGCPFCADDSPCFVQEDKYLRLAI 420  
QY 421 ISFOGLCMLLDFVSMVYHFKAKSIRASGLILLETILFGSLILYFPVILYFEPSTFR 480  
DB 421 ISFOGLCMLLDFVSMVYHFKAKSIRASGLILLETILFGSLILYFPVILYFEPSTFR 480  
QY 481 CILLRWARLLGFATVYGTVTLKHLRVKLVLSRTAQRIPYMTGGRRWMLAVILLVVEWF 540  
DB 481 CILLRWARLLGFATVYGTVTLKHLRVKLVLSRTAQRIPYMTGGRRWMLAVILLVVEWF 540  
QY 541 LIGWTSVVCQNLKQISLIGQKTSDDLIFNMCNLDIDRWDMYTAGVMSVSYDGLTIFQ 599  
DB 541 LIGWTSVVCQNLKQISLIGQKTSDDLIFNMCNLDIDRWDMYTAGVMSVSYDGLTIFQ 599  
RESULT 2  
AAE06642  
ID AAE06642 standard; Protein; 1215 AA.

XX  
AC AAE06642;  
XX  
DT 16-OCT-2001 (first entry)  
XX  
DE Human G-protein coupled receptor (NGPCR) #1.  
XX  
KW Human; G-protein coupled receptor; NGPCR; gene therapy; cancer; diabetes;  
KW inflammation; immune disorder; heart disease; obesity; coronary disease;  
KW metabolic disorder; physiological disorder; therapeutic; drug screening;  
KW signal transduction; behavioural disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200157086-A2.  
XX  
XX 09-AUG-2001.  
XX  
PF 01-FEB-2001; 2001WO-US03648.  
XX  
PR 04-FEB-2000; 2000US-0180414.  
XX  
PA (LEXI-) LEXICON GENETICS INC.  
XX  
PI Donoho G, Hilbun E;  
XX  
XX WPI; 2001-488870/53.  
DR N-PSDB; AAD12292.  
XX  
PT Novel isolated polynucleotides encoding human G protein coupled  
PT receptor (NGPCR), useful as probe and for treating disease involving  
PT GPCR, detecting mutant NGPCR or abnormally expressed NGPCRs for  
PT diagnosing disease  
XX  
PS Claim 2; Page 66-69; 73pp; English.  
XX  
CC The present sequence is human G-protein coupled receptor (NGPCR) which  
CC is a 7 transmembrane protein of G-protein coupled receptor (GPCR) super  
CC family. NGPCR is expressed in hypothalamus, foetal brain, brain,  
CC cerebellum and testis. NGPCR spans the cellular membrane and is involved  
CC in signal transduction pathways. Human NGPCR is useful for treating  
CC obesity, inflammation, immune disorders, diabetes, heart and coronary  
CC disease, metabolic disorders and cancer. NGPCR's are also used as  
CC reagents in assays for screening compounds that are useful in the  
CC therapeutic treatment of physiological and behavioural disorders. NGPCR  
CC is useful for identifying compounds that modulate NGPCR gene expression  
CC and/or gene product activity. NGPCR DNA is also useful for the diagnostic  
CC and prognostic evaluation of disorders related to NGPCR function.  
CC Nucleotide constructs encoding functional NGPCR and mutant NGPCR variants  
CC are used in gene therapy. NGPCR probes are also useful for the  
CC identification, selection and validation of novel molecular targets for  
CC drug discovery.  
XX  
XX Sequence 1215 AA;

Query Match 97.5%; Score 3119.5; DB 22; Length 1215;  
Best Local Similarity 98.5%; Pred. No. 1.3e-291;  
Matches 585; Conservative 2; Mismatches 2; Indels 5; Gaps 1;  
QY 1 MGAMAYPLLCLLLAQLGLGAVGASRDQGRDPSRPTPKGPKHAQPGGRASDSSAP 60  
DB 1 MGAMAYPLLCLLLAQLGLGAVGASRDQGRDPSRPTPKGPKHAQPGGRASDSSAP 60  
QY 61 WSRSTGDTTLAOKLAEEVPMVASYLTGDSHQLKRANCSGRYELAGLPKWPALASAH 120  
DB 61 WSRSTGDTTLAOKLAEEVPMVASYLTGDSHQLKRANCSGRYELAGLPKWPALASAH 120  
QY 121 SLHRALDTLTATNLFNVLNMQSNKREONLQDDLDWYQALVMSLLEGEPSISRAAITFST 180  
DB 121 SLHRALDTLTATNLFNVLNMQSNKREONLQDDLDWYQALVMSLLEGEPSISRAAITFST 180  
QY 181 DLSAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPPLHRRGN 240  
DB 181 DLSAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPPLHRRGN 240

Db 181 DLSAPAPQVFLQATREESRILLQDLSAPHLANATLETWFHGLRRKWRPHLHRRGPN 240  
QY 241 QGPRGLGHSWRRKDGKSHFKSPPYLECENGSYKPGWLVTLSAIYGLQPNLVPEF 300  
Db 241 QGPRGLGHSWRRKDGKSHFKSPPYLECENGSYKPGWLVTLSAIYGLQPNLVPEF 300  
QY 301 RGVKVDINLQKVDIDQCSDDGWFSTHCKHLNNSCMPKIGLGFVLGAYECICKAGFYH 360  
Db 301 RGVKVDINLQKVDIDQCSDDGWFSTHCKHLNNSCMPKIGLGFVLGAYECICKAGFYH 360  
QY 361 PGVLPVNNFRRGPDQHISGSKDVSEAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420  
Db 361 PGVLPVNNFRRGPDQHISGSKDVSEAYVCLPCREGCPFCADDSPCFVQEDKYLRLAI 420  
QY 421 ISFGCLMLLDFVSMVLVYHFRKAKSIRASGLILLETILFGSLLLYFPVVILYFEPSTER 480  
Db 421 ISFGCLMLLDFVSMVLVYHFRKAKSIRASGLILLETILFGSLLLYFPVVILYFEPSTER 480  
QY 481 CILLRWARLLGFATVYGTVTCLKLHRLVKVFLSRTAQRIPYMTGGRVMRLAVILLVFWWF 540  
Db 481 CILLRWARLLGFATVYGTVTCLKLHRLVKVFLSRTAQRIPYMTGGRVMRLAVILLVFWWF 540  
541 LGWTSSVQCNLEKQISLIGOGKTSQHLIFNMCLIDRWDMYMTAVG-----MWSL 589  
Db 541 LGWTSSVQCNLEKQISLIGOGKTSQHLIFNMCLIDRWDMYMTAVG-----MWSL 589  
541 LGWTSSVQCNLEKQISLIGOGKTSQHLIFNMCLIDRWDMYMTAVG-----MWSL 594  
RESULT 3  
ABB70858  
ID ABB70858 standard; Protein; 669 AA.  
XX AC ABB70858;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 39366.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PR 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
XX N-PSDB; ABL14961.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX Disclosure; SEQ ID NO 39366; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX (ABB57737-ABB72072).  
XX The sequence data for this patent did not form part of the printed  
PI

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 669 AA;  
Query Match 10.6%; Score 339; DB 22; Length 669;  
Best Local Similarity 25.6%; Pred. No. 2.4e-23;  
Matches 127; Conservative 71; Mismatches 200; Indels 98; Gaps 22;  
QY 110 GKWPALASAPSLHRAIDLTLTATNFLNVLNMQSNKSRQNLQDDLDLWYQALVMSLLEGE 169  
Db 6 GWPC-RMFYDIDIAEDAROFIEFL-----SGKFPNANTPIAID-----EP 47  
QY 170 SISRAITFSTDSLAPAPQVFLQATREESRILLQDLSAPHLA-----NATLETWF 223  
Db 48 --TRAEVSRANGIAS-----YALNEDDNLLAFAIPAISIHTVVVFRDNVTIPDQV 98  
QY 224 HG---LRRKWRPHLHRRGPNQGPRLGHSWRRKDGKSHFKWSPPYLECENGSYKPG 280  
Db 99 HNKAYLGSYWR-----LGAANSTDGTO-----EWGAPFRDCNLLTRRWL 139  
QY 281 WLVTLSAIYGLQPNLVPEFPGVMKVDINLQKVDIDQCSDDGW---FSGTHKHLNSEC 337  
Db 140 WPFRTSFS-----EHR-IKVVAAPFAADEVC-NDGLEEVFGRRHGCDRNTTFC 187  
QY 338 MPKIGL-GEVLGAYECICKAGFYHGPVLPVNNFRRGPDQHISGSKDVSE--EAYVCLP 394  
Db 188 LLTENKPAATRDVYTCLCRESYLPN-STLOGFR-----GDRVELSEGVDNYSIP 237  
QY 395 CREGPCFCAADSPC---FVQE-----DKYLRLAIISFGCLMLLDFVSMVLVYHFRKAKSI 447  
Db 238 CPQCTNCDNSGVCLTFQEEVLNVDACLRLVAIVLGACILCCIVLGVIVFRQRCKAI 297  
QY 448 RASGLILLETILFGSLLLYFPVVILYFEPSTERCILLRWARLLGFATVYGTVTCLKLHRLV 507  
Db 298 ASGMWTVLETILGIVLLYASVAVHFFPASTERCLLEPWLRELGFTICGAILLKLYRHL 357  
QY 508 KVLSTAQRIPIYMTGGRVMRLAVILLVFWFLIGTSSVQCNLEK-QISLIGOGKTS 566  
Db 358 VDFRTRKAHRW-VLRDVLKLYGTWVFAVICYMAAFTASSLDLLESAQLESUREADT-- 414  
QY 567 HLIFNMCLIDRWDMYT 582  
Db 415 ----NTCHPLKWELVT 426  
RESULT 4  
AAU38983  
ID AAU38983 standard; Protein; 669 AA.  
XX AC AAU38983;  
XX DT 16-JAN-2002 (first entry)  
XX DE Drosophila G-protein coupled receptor, GCPR #61.  
XX KW Drosophila; G-protein coupled receptor; GCPR; insecticide; diagnostic;  
XX KW mutation detection.  
XX OS Drosophila melanogaster.  
XX PN WO200170980-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09341.  
XX PR 23-MAR-2000; 2000US-191638P.  
XX PR 18-JUL-2000; 2000US-0618893.  
XX (PEKE ) PE CORP NY.  
XX Cravchik A;  
PI

XX WPI: 2001-616405/71.  
 DR N-PSDB; AAS57192, AAS57193.  
 XX Sixty six Drosophila Melanogaster G-protein coupled receptors (GPCR),  
 PT useful in the treatment and diagnosis of GPCR-related conditions and  
 PT for identifying GPCR modulators for use as insecticides -  
 XX  
 XX Claim 1; Page 275-277; 392pp; English.  
 XX  
 XX The invention relates to sixty six novel isolated Drosophila  
 CC melanogaster G-protein coupled receptors (GPCR). The GPCR proteins and  
 CC nucleic acids are useful in the treatment and diagnosis of GPCR-related  
 CC conditions. The GPCR proteins and nucleic acids are also useful for  
 CC identifying modulators of GPCR proteins for use as insecticides. The  
 CC nucleic acid can also be used to detect mutations in GPCR genes and gene  
 CC expression products such as mRNA. AAU38923-AAU38988 represent D.  
 CC melanogaster G-coupled protein receptor amino acid sequences of  
 CC the invention.  
 SQ Sequence 669 AA;  
 Query Match 10.6%; Score 339; DB 22; Length 669;  
 Best Local Similarity 25.6%; Pred. No. 2.4e-23;  
 Matches 127; Conservative 71; Mismatches 200; Indels 98; Gaps 22;  
 QY 110 GKWPALASAPSLHRALDILTHATNPLNVMLOSKREONLODDLDYQALVWSLLEGEP 169  
 DB 6 GKWPC-RMFYDYTDIADAARFIEFL-----SGKFPNANTPAID-----EP 47  
 QY 170 SISRAAITSTSLSPAPQVFLQATREESRILLQDLSSAPHLA-----NATLETWF 223  
 DB 48 --TRAESVRANGIAS-----YALNEDDNLFAFAIAPSIHTVVVFRDNVTIPDQV 98  
 QY 224 HG---LRRKWRPHLHRGNQGRGLGHSWRKDKGLGKSHFKNSPPYLECENGSKPG 280  
 DB 99 HNKAYLGSYWRE-----LGAAWNSTDGTQ-----EWGAPFRDCNLLTRRWL 139  
 QY 281 WLVTLSAAYGLQPNLVPFRGVKMDINLQKVIDQCSDDGW---FSGTHKCHLNSEC 337  
 DB 140 WPRISFS-----EHR-IKVVAFAFIADDEVC--NDGLEEVFGRHGGDRNTTC 187  
 QY 338 MPKIGL-GEVLGAYEICKAGEYHPGVLPVNNFRRRGPQDHISGSKDVSE--EAYVCLP 394  
 DB 188 LLTENKPAATRDVYTCLCRESYLPN-STLOGFR-----GDRVELSEGVDNYSVIP 237  
 QY 395 CREGCPFCADDSPC--FVQE-----DKYLRALISFGQLCMLLDVFSMLVYVHFRKAKSI 447  
 DB 238 CPGGCTNCDNSNGVCLTFQEEVLNVNDACLRLLVAIVLGACILCCIVLGVIVFRQKCKAI 297  
 QY 448 RASGLILLETILFGSLLLYFPVIVLYFEPSTFCILLRWALLGFATVYGTVTLKLRHYL 507  
 DB 298 ASGMWTVLETILGIVLLVAVSNAVHFFASTRECLLEPWRLGLFTCYGAILIKDYRL 357  
 QY 508 KVFLSRTAQIPYMTGGRVMRLAVILLVFWFLIGTSSVQCNLEK-QISILGQGTSD 566  
 DB 358 VDFTRKAHRW-VLRDVLKLVGTWVFVAVICYMAAFTASSLDLLESAQLESIREADT-- 414  
 QY 567 HLIFNCLIDRWYMT 582  
 DB 415 ----NTCHPLKWLVT 426  
 RESULT 5  
 ABB67185  
 ID ABB67185 standard; Protein; 176 AA.  
 XX  
 XX ABB67185;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT Drosophila melanogaster polypeptide SEQ ID NO 28347.  
 DE Drosophila melanogaster.  
 XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX Drosophila melanogaster.  
 OS  
 PN WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR  
 XX 11-JUL-2000; 2000US-0614150.  
 PN  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI WPI: 2001-656860/75.  
 XX N-PSDB; ABL11288.  
 DR  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PS Disclosure; SEQ ID NO 28347; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 176 AA;  
 SQ  
 Query Match 7.5%; Score 238.5; DB 22; Length 176;  
 Best Local Similarity 31.5%; Pred. No. 1.6e-14;  
 Matches 52; Conservative 25; Mismatches 55; Indels 33; Gaps 4;  
 QY 301 RGVKVDINLQKVIDQCSDD-----GWFSGTHKCHLNSECMPKIGLGVLGAYE 351  
 DB 9 RGTSGIDILDRVIDDQCPQRHTPGTKRPLNIFAGTDKCKQRTMCEAIMGLGFRGYSK 68  
 QY 352 CICKAGFYHPGVLPVNNFRRRGPQDHISGSKDV-----SEAYVCLPCRE 397  
 DB 69 CLCRKGFYFPDIVSQHKF-----FNGSLLEEEYEKMLGKNSTYNSNSEYCLPCAE 120  
 QY 398 GPCFCADDSPCFQVEDKYLRALISFGQLCMLLDVFSMLVYVHFR 442  
 DB 121 GDCSDSDSPCIAALNWPMTSILAL--ACIVIGLLPPAAWFTFR 163  
 RESULT 6  
 ABB65498  
 ID ABB65498 standard; Protein; 976 AA.  
 XX  
 XX ABB65498;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT Drosophila melanogaster polypeptide SEQ ID NO 23286.  
 DE Drosophila melanogaster.  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX Drosophila melanogaster.  
 OS  
 XX





Db 265 SFINKYRREHPGKLAERVVSVIRKMSVEEKQAFQRIQTGTTHVHFEEVEVPVPSND 324  
 QY 344 -----GFVLGAYECICKAGFYHPGVLPVNNFR 371  
 Db 325 VIDIVECDYRQPSAETPTVTGKLLTLTGSSWTRGSCICLORGFY-----SL 374  
 QY 372 RGPD-----OHISGSTKDYSEAYVCLPCRGCPFCADSPCFVQEDKYIRLAI 420  
 Db 375 RHPDGFNGTIMEAWQEQDNISNYSEVFKCLPCAPGCDTCTGPEPCLANYHWPFRISL 434  
 QY 421 ISFQGLCMLLDFVSMVYVHFRKASIRASGLILLETILFGSLLLY 466  
 Db 435 LTISIGCAGCTEVLAGYLFRRHRVVKVASPIFLMITLIGCAIMY 480

RESULT 8  
 AAY45000  
 AAY45000 standard; Protein; 974 AA.

AAY45000;

20-JUN-2000 (first entry)

Human calcium sensing receptor isoform, Casrd.

Calcium Sensing Receptor; Casrd; isoform: human; splice variant;  
 KW hypotensive; osteopathic; receptor activity; calcium level; modulator;  
 KW treatment; hyperparathyroidism; osteoporosis; Paget's disease;  
 KW hypercalcaemia malignancy; hypertension; gene therapy.

OS Homo sapiens.  
 OS Synthetic.

PN WO200006601-A1.

XX 10-FEB-2000.

XX 28-JUL-1999; 99WO-US17116.

XX 30-JUL-1998; 98US-0094702.

XX (AVET) AVENTIS PHARM PROD INC.

PI Yu KT, Labaudiniere RF, Thrower LW;

XX WPI; 2000-195263/17.

XX N-PSDB; AAZ50616.

PI Nucleic acids encoding isoforms of human calcium sensing receptor for  
 PT treating, e.g. hyperparathyroidism or osteoporosis -

XX Claim 24; Page 66-69; 81pp; English.

XX The present amino acid sequence is an isoform of the human calcium  
 CC sensing receptor Casrd, expressed in the kidney. It is a splice variant  
 CC of the wild type Casr, that arise from deletion of amino acids  
 CC 361-465, from the extracellular domain, that comprise acidic residues.  
 CC It has hypotensive and osteopathic activity. The Casr isoforms can be  
 CC used to identify agonists and antagonists that modulate the receptor  
 CC activity and calcium levels. These modulators are useful for treating  
 CC hyperparathyroidism or osteoporosis, Paget's disease, hypercalcaemia  
 CC malignancy or hypertension. The DNA sequence is also useful for altering  
 CC the Casr activity and in gene therapy.  
 CC Note: This sequence has deletion of amino acids 361-465, according to  
 CC the sequence shown in the specification, but has been stated as  
 CC 358-462 in the claims.

XX Sequence 974 AA;

Query Match 4.4%; Score 141.5; DB 21; Length 974;

Best Local Similarity 19.08; Pred. No. 0.00053;

Matches 114; Conservative 89; Mismatches 246; Indels 151; Gaps 26;

QY 75 ABEVPMVDVASYLYTGDSHQLKRANGSGRYELAGLPCKWPAALASHPSLHRLDITLTHATN 134  
 Db 132 SEHIFSTIAVVGATGSGYSTAVANLLGLFYI-----POVSYASS--RLLSNKNQPKS 182  
 QY 135 FUNVMLOSKREONLQDDLDWYQALVWSLL-----EGEPSISRAAITFSTDSLSAP 186  
 Db 183 FLRT-IPNDEHQATAMADIIIEYF--WNWVGTTAAADDDYGRPGIEKFPREEAERDICTD 238  
 QY 187 APOVFLQATREESRLLQDLSASSAPHLANATLETWFLHCLRRKWRPHLHRRCPNPGP--- 243  
 Db 239 FSELISQYSDEE-----EIOHVEVIQNSTAKV-----IVVFSSGPDLEPLIK 281  
 QY 244 ----RGL-GHSWRRKDGGLGDKSHFKWS-----PPYLEGECNGSYKPGWLVTLSAIYG 291  
 Db 282 EIVRNITGKIWASEA-----WASSSLIAMPQYFHVVGTTIG-----FA 321  
 QY 292 LQPNLVPEFRGVMKYVDINLQKVIDQSSDGFWSGTHKCHL-----NN-----SECM 338  
 Db 322 LKAGQIPGFRFLK-KVHPRKSVHNGFAKEFW-EETFNCHLRHLNFTNNMGEQVTFDECG 379  
 QY 339 PIKGLGFVL-----GAYECICKAG---FYHPGVLPVNNFRRRGPDQH-- 377  
 Db 380 DLVGNYSIINMHLSPEDGSIVFKEVGYNVYAKGERLFINEEKLWSGFSREVFPFNSCS 439  
 QY 378 ---ISGSTKDYSEE---AYVCLPCREG-----CPTCADD-----SPCFVOED 413  
 Db 440 RDLAGTRKGIIEGPTCCFCEVCEGDEYDETDASACNKPDDFWSENNHTSCIAKEI 499  
 QY 414 KYLR-----LAIISFQGLCMLLDFVSMVYVHFRKASIRASGLILLETILFGSLLLYF 467  
 Db 500 EFLSWTEPFGIALTFLFAVLGIFLTAFLGVFIKPRNTPIVKATNRELSYLLFLSLCCFS 559  
 QY 468 PVILYFEPSTFRCILLRWARLLGFATVYGVTVLKLHVLFVLSRTAQRIPYMTGGRMV 527  
 Db 560 SSLFFIGEPODMTCRLRQPAFGISFVLCISCLIVKTNVLLVFEAKIPTSFHRKMWGLNL 619  
 QY 528 RMLAV-----ILLVFVFWFLIGWTSV-CONLEKQISLI-----GOGKTSDDLIFNMCLI 575  
 Db 620 QLLVFLCTFMQIVICVILWLYTAPPSSYRNQOELEIIFITCHEGSLMALGFLIGYTCLL 679

RESULT 9

AAY49126

ID AAY49126 standard; Protein; 917 AA.

XX AAY49126;

XX 07-JAN-2000 (first entry)

XX Chimeric hCAR/hmGluR2 protein sequence.

XX G-protein fusion receptor; CaR; calcium receptor; GluR; head injury;  
 KW metabotropic glutamate receptor; GABABR; chimeric receptor; stroke;  
 KW gamma-aminobutyric acid receptor; allosteric modulator; antagonist;  
 KW spinal cord injury; epilepsy; ischaemia; hypoglycaemia; anoxia;  
 KW Alzheimer's disease; hyperparathyroidism; osteoporosis; depression;  
 KW cognitive disorder.

XX Homo sapiens.

XX WO9951641-A1.

XX 14-OCT-1999.

XX 02-APR-1999; 99WO-US07333.

XX 03-APR-1998; 98US-0080671.

XX (NPSP-) NPS PHARM INC.

XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;  
 PI Simin RT;

XX



KW cognitive disorder.  
 XX Homo sapiens.  
 OS  
 XX WO9951641-A1.  
 PN  
 XX 14-OCT-1999.  
 PD  
 XX  
 XX  
 XX 02-APR-1999; 99WO-US07333.  
 PF  
 XX  
 XX 03-APR-1998; 98US-0080671.  
 PR  
 XX  
 XX (NPS- ) NPS PHARM INC.  
 PA  
 XX Stormann TM, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;  
 PI Simin RT;  
 PI  
 XX WPI; 1999-610995/52.  
 DB N-PSDB; AA231058.  
 PT  
 PT New G-protein fusion receptors and chimeras containing domains from  
 PT different receptors, used to screen for modulators, potentially useful  
 PT e.g. for treating or preventing stroke or Alzheimer's disease -  
 XX  
 XX Example 1; Fig 12; 255pp; English.  
 PS  
 XX The invention relates to G-protein fusion receptors (1) comprising:  
 CC (1) in the N to C direction, extracellular (ECD), transmembrane (TMD) and  
 CC intracellular (ICD) domains, each chosen independently from a Car  
 CC (calcium receptor), GluR (metabotropic glutamate receptor) and GABABR  
 CC (gamma-aminobutyric acid receptor); (2) an optional linker attached to  
 CC the C-terminus of ICD; and (3) a G-protein (GP) linked to ICD or the  
 CC linker. (1), and recombinant chimeric receptors (CR) without the GP  
 CC component, are used to assess function of the various domains and to  
 CC identify compounds (e.g. allosteric modulators or antagonists) that act  
 CC on these domains. The modulators are potentially useful for treating or  
 CC preventing diseases associated with the receptors, e.g. stroke, head or  
 CC spinal cord injury, epilepsy, ischaemia, hypoglycaemia, anoxia,  
 CC Alzheimer's disease, hyperparathyroidism, osteoporosis, cognitive  
 CC disorders and depression. Nucleic acid (II) that encodes (I) is used:  
 CC (1) for recombinant production of corresponding proteins; and (2) to  
 CC produce cells used in screening for modulators. Use of Car and mGluR  
 CC domains allows presentation of GABABR domains, to a binding agent, in a  
 CC form more like the natural domain structure compared with use of  
 CC incomplete receptors, lacking one or more domains. By shuffling different  
 CC domains, agents can be identified that affect particular domains of a  
 CC receptor.  
 SQ Sequence 1276 AA;

## RESULT 12

AAO15093  
 ID AAO15093 standard; Protein; 1276 AA.  
 XX  
 XX AAO15093;

DT 22-AUG-2002 (first entry)  
 DE  
 XX Human phCar/hmGluR2\*Gqi5 fusion construct protein.

XX Human: G-protein fusion receptor; extracellular domain;  
 KW transmembrane domain; intracellular domain; Car; mGluR; GABABR;  
 KW modulator identification.  
 XX  
 XX Chimeric - Homo sapiens.  
 XX  
 XX WO200229033-A2.  
 PN  
 XX 11-APR-2002.  
 PD  
 XX 03-OCT-2001; 2001WO-US31074.  
 PF  
 XX 03-OCT-2000; 2000US-0679664.  
 PR  
 XX (NPS- ) NPS PHARM INC.

XX Stormann T, Hammerland LG, Storjohann LL, Busby JG, Garrett JE;  
 PI Simin RT;

XX WPI; 2002-330170/36.

XX Novel G-protein fusion receptor, useful for identifying modulators of  
 PT Car, mGluR and GABABR, comprises G-protein joined to the intracellular  
 PT domain of the receptor -  
 XX  
 XX Example 3; Fig 12; 168pp; English.

CC The invention comprises G-protein fusion receptors - comprising  
 CC extracellular, transmembrane and intracellular domains similar to Car,  
 CC mGluR or GABABR receptor sequences. The G-protein fusion receptors of the  
 CC invention may also possess a linker joined to the carboxy terminus of the  
 CC intracellular domain, and a G-protein joined to the linker. The G-protein  
 CC fusion receptors of the invention are useful for identifying modulators  
 CC of Car, mGluR and GABABR for use in treating associated conditions. The  
 CC present amino acid sequence was used in the production of the invention.

SQ Sequence 1276 AA;

Query Match 4.4%; Score 139.5; DB 23; Length 1276;  
 Best Local Similarity 24.2%; Pred. No. 0.0013;  
 Matches 68; Conservative 43; Mismatches 111; Indels 59; Gaps 15;

QY 311 OKVIDQCSSDGFSGTHKC---HLNNE-CMPIKGLGFLGAYECICKAG---FYHPGV 363  
 DB 475 EQVTFDEC---GDLVGNYSIINWHLSPEDGSIYFKEVGY---YNYAKKGERLFINEEK 527

QY 364 LPVNNFRRRGPDQH-----ISGSTKDVSEE-----AYVCLPCREG-----CPFCAD 404  
 DB 528 ILWSGFSREVFPFSGNRDCLAGTRKGIIEGPTCCFCEVCPDGEYSDETDASACNKKCPD 587

QY 405 D-----SPCFVQEDKYLK-----LAIISFGGLCMLLDFVSMVLYVHFRKAKSIR 448  
 DB 588 DFWSNENHSTCFELPQEIYRWGDWAVGPVTIACLGALATL--FVLGVFVRH-NATPVVK 644

QY 449 ASGLILLETILFGSLLLYFPVIVLYFEPSTFRCILLRWARLLGFATVYGTVTLKLRVLYK 508  
 DB 645 ASGRELCYILGGVFLCYCMTFIFTAKSTAVCTLRRLGLGTAFSVCYSALLTKNRIAR 704

QY 509 VF--LSRTAQRIPYMT--GGRVNRMLAVI-----LLVVFWEFLI 542  
 DB 705 IFGGAREGAQRPRFTSPASQVAICLALISGQLLIIVWALVV 745



CC veterinary medicine; and (ii) genes that encode polypeptides involved in  
 CC assembly of functionally related GABA-B receptors in insects. This  
 CC sequence represents a fruitfly (Drosophila melanogaster) GABA-B receptor  
 CC which is described in the method of the invention.

XX  
 SQ Sequence 1305 AA;  
 Query Match 4.2%; Score 134.5; DB 22; Length 1305;  
 Best Local Similarity 18.8%; Pred. No. 0.004;  
 Matches 106; Conservative 80; Mismatches 179; Indels 199; Gaps 27;  
 QY 63 RSTDGTLAQLAEVPMVDVASYLYTGDHQLKANCGR-----YELAGLFGKWP--A 114  
 DB 344 RETDRIIIGSFQELAPOLCEAY-----RLRMFGADYAWILHESGAPW-WPDQR 394  
 QY 115 LASAFLSHRALDTL-----TH-----ATNPLNVLQSNKSRQNLQDLDLWYQALVW 162  
 DB 395 TACSNHELQAVENLIVVSTHNSIVGNVSYGLNHHMFNSQLRKQSAQ----- 443  
 QY 163 SLLEGPSISRAAITFSTDSLAPAPQVFLQATREESR-----ILLQDLSS 208  
 DB 444 --PHGODGFG-----SCYGRISATQSDSRRRRRRGVGTSGGHLFPEAISQ 490  
 QY 209 SAPHLANATLETWFHGLRKRKPHLHRRGPNQPRGLGHSWRRKDLGGDKSHFKWSP 268  
 DB 491 YAPQ-----TYDAVWATALALRAAEHWRNEQS-----KLDGFDYTRSDMAWE-- 535  
 QY 269 YLE-----CENGSKPKPWLTLSAIYGLQPNLVPFRGVMKVDINLOKVD 314  
 DB 536 FLQGMKHLFLGVSGVPVSFGDRVG-----TFATFYQIQ-----RGL-----LEPVA 577  
 QY 315 IDQCSDDGWFSTGHKHLNNSCMPKIGLGLGAYECICKAGFYHPGVLPVNN--FRRR 372  
 DB 578 L-----YYPATDALDFRCPCRPVK-----WHSQVPTAKRVFKLR 613  
 QY 373 GPDQHTSGTKDVEEAYVCLPCREGCPFCADDSPCFVQEDKYRLAIISFOGLCMLDF 432  
 DB 614 -----VATIAPL-----AFYTIATLSSVGIATIAIF 639  
 QY 433 VSMVLVYHFRKAKSIRASGLILETLFGSLLLYFPVVILYFEPN-----FRCILLR 485  
 DB 640 LAFNL--HERKKAIKLSPKLSNTAVGICFYATVILLGLDHSFLPSAEDSFATVCTA 697  
 QY 486 WARLL--GFATVYGTVTLKHLRVKLVFLSRTAQ-----RIPYMTGGRVMRLAVI 533  
 DB 698 RYVLLSAGSLAAGSMFATYRVHRIF--TRTGSVFKDKMLQDILILLVGG---LLLVD 753  
 QY 534 LLVWFELIGWTSSVCQNLKQIS 557  
 DB 754 LLVTLWVTDPMERHLHNLTLAIS 777

RESULT 15  
 AAW32059  
 ID AAW32059 standard; Protein; 1026 AA.  
 XX  
 AC AAW32059;  
 XX  
 DT 27-MAR-1998 (first entry)  
 XX  
 DE Dogfish shark kidney calcium receptor related protein (SKCar-RP).  
 XX  
 KW Calcium receptor related protein; Car-RP; dogfish shark; SKCar-RP;  
 KW polycation-sensing receptor; aquaculture; fish farming;  
 KW salinity tolerance.  
 XX  
 OS Squalus acanthias.  
 XX  
 FH Location/Qualifiers  
 FT 351..395  
 FT Region  
 FT /note= "region in extracellular domain that is  
 highly divergent from mammalian pVCR"  
 FT 870  
 FT Region

FT  
 FT  
 XX  
 PN W09735977-Al.  
 XX  
 PD 02-OCT-1997.  
 XX  
 PF 27-MAR-1997; 97WO-US05031.  
 XX  
 PR 27-MAR-1996; 96US-0622738..  
 XX  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 XX  
 PI Brown E, Harris HW, Hebert S;  
 XX  
 DR WPI; 1997-489640/45.  
 DR N-PSDB; AAT89290.  
 XX  
 PT New isolated Aquatic polyvalent cation-sensing receptor - used to  
 develop products for increasing or decreasing the salinity tolerance  
 of fish for use in aquaculture  
 PT  
 PS Claim 9; Fig 5A-B; 57pp; English.  
 XX  
 CC This protein comprises dogfish shark kidney calcium receptor  
 related protein (SKCar-RP), an aquatic polyvalent cation-sensing  
 receptor (pVCR). Its amino acid sequence was deduced from a kidney  
 cDNA clone (see AAT89290). It shows 74% homology to rat kidney  
 pVCR and bovine parathyroid pVCR and possesses general features  
 that are homologous to pVCR proteins, including a large  
 extracellular domain, 7 transmembrane domains and a cytoplasmic  
 C-terminal domain. Also claimed are: an isolated pVCR present in  
 the plasma membranes of aquatic species, especially on the apical  
 membrane of epithelial cells of elasmobranch fish, particularly  
 from cells found in the collecting duct or late distal tubule in  
 the kidney, intestine, gill, rectal gland, gonad or brain; an  
 antibody that specifically binds to a pVCR; and a method of  
 screening for aquatic pVCR agonists and antagonists. Modulation of  
 the expression of the aquatic pVCR activates or inhibits aquatic  
 pVCR mediated ion transport and endocrine changes that permit fish  
 to adapt to fresh or salt water. The method facilitates the  
 aquaculture of marine fish and can provide for the development of  
 marine fish that are easily adaptable to fresh water aquaculture.

SQ Sequence 1026 AA;  
 Query Match 4.2%; Score 133.5; DB 18; Length 1026;  
 Best Local Similarity 19.5%; Pred. No. 0.0034;  
 Matches 91; Conservative 59; Mismatches 167; Indels 149; Gaps 18;  
 QY 234 LHR-----RGPNQGRGLGHSWR---RKDGL---GGDKSHFKWSPYLCENG 276  
 DB 363 LHREDLTOLKNSKVPSPHGAAGGSGKAGNSRRTALRHPTGTEENITSVETPYLDITHL 422  
 QY 277 YKPGWLVTLSAIYGLQPNLVPFRGVKMDINLOKVIDOCSS----- 320  
 DB 423 ISYNNYVAVYSIAHALQ-----DIHCKPGTGIFANGSCADIKV 462  
 QY 321 DGWFGTHKHLNNSECM-----PIRG-----LGFVLGAYECIC 354  
 DB 463 EAWQVLNHLHLKFTNSMGEQVDFDDQDGLKGNFTIINQLSAEDSVLFHEGVNAYA 522  
 QY 355 K-----AGFYHCVLPVNNFRRRGPPQDHISGSKDYSEE-----AYVCLPCR 396  
 DB 523 RPSDRNLINEXKILWSGF--SKWPFNSCSR---DCVPGTRKRIIEGTEPTCCFECMAC 576  
 QY 397 EG-----CPFCADD-----SPCFVQEDKYLR-----LAIISFQGLCMLDFV 433  
 DB 577 EGEFSDENDASACTKCPNDFWSNENHTSCIAKEIYLSWTEPFGLIATIFAVGLIITSF 636  
 QY 434 SMLVYHFRKAKSIRASGLILETLFGSLLLYFPVVILYFEPSTFRICLLRWALLQFA 493  
 DB 637 VLGVFIKFRNTPIVKATNRELSYLLLSLFCFSSSLIFIGEPDWTCLRQPAFGISFV 696

/note= "region in C-terminal domain that is  
 highly divergent from mammalian pVCR"

QY 494 TVYGTVTLKHLRVKFLSRTAQIPYMTGGRVMRLAVILV-----VFWFLIGWTS 546  
Db : : | | | | : : : : | : : : : :  
697 LCISCILYKTNRVLLVFEAKIPTSLHRKWVGLNLQFLLVFLCIIQIVTCIIWLYTAPPS 756  
QY 547 SVCQMLEXQOISLIGQKTSDDLIFNMCLIDRDYMTAVGMMSLVSY 592  
Db : : | : : : : : : : : : : : : : : :  
757 SY-RNHELE-----DEVIFITC-----DEGSLMALGFLIGY 786

Search completed: February 15, 2003, 08:29:00  
Job time : 38.7095 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2003, 05:55:14 ; Search time 16.1803 Seconds  
(without alignments)  
3558.936 Million cell updates/sec

Title: US-09-775-181-4

Perfect score: 3199

Sequence: 1 MGAWAYPLLLCLLQAQLGLG.....YMTAVGMWSLVSDGLTIFQ 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	4.3	999	2 T27628	hypothetical prote
2	133.5	4.2	1079	2 I59362	calcium/polyvalent
3	127	4.0	1078	2 A56715	calcium receptor (
4	123	3.8	1085	2 S40476	Ca(2+)-sensing rec
5	120.5	3.8	879	2 JC7160	metabotropic gluta
6	119.5	3.7	1088	2 B56715	calcium receptor (
7	116	3.6	872	2 JH0561	metabotropic gluta
8	115	3.6	1199	2 A41939	G protein-coupled
9	114	3.6	1180	2 JC2132	metabotropic gluta
10	114	3.6	1212	2 JC2131	metabotropic gluta
11	112	3.5	879	2 JH0562	metabotropic gluta
12	111	3.5	1171	2 A42916	metabotropic gluta
13	110.5	3.5	3507	2 T34513	hypothetical prote
14	109.5	3.4	868	2 JC5701	ErbB kinase activa
15	107.5	3.4	2437	2 S53611	MIBP1 protein - ra
16	106.5	3.3	850	2 JC5700	ErbB kinase activa
17	106.5	3.3	860	2 JC5702	ErbB kinase activa
18	106	3.3	383	2 S53716	delta-like homeoti
19	106	3.3	551	2 T30806	metabotropic gluta
20	104.5	3.3	644	1 A40212	uromodulin precurs
21	104.5	3.3	644	2 I84634	Tamm-Horsfall prot
22	104.5	3.3	956	2 A57121	thrombospondin 3 p
23	100.5	3.1	1218	2 S71376	glutamate receptor
24	100	3.1	1993	2 T30902	sodium channel SCA
25	99	3.1	669	2 I38029	matrix metalloprot
26	98.5	3.1	459	2 JC5139	vitronectin precur
27	98	3.1	2531	2 T31070	notch homolog - se
28	98	3.1	2871	2 A55567	fibrillin I - bovi
29	97.5	3.0	385	2 A54785	preadipocyte facto

30 97.5 3.0 3002 2 A47221 fibrillin 1 precur  
31 97 3.0 862 2 H82182 conserved hypothet  
32 97 3.0 2918 2 A54105 fibrillin-2 precur  
33 96.5 3.0 723 1 VCPVPP coat protein vpi -  
34 96.5 3.0 986 2 G65116 hypothetical prote  
35 96.5 3.0 1203 2 A49175 Motch B protein -  
36 96.5 3.0 2871 2 A55624 fibrillin-1 precur  
37 95.5 3.0 858 2 JC7683 taste receptor T1R  
38 95 3.0 468 2 T12725 terminase large ch  
39 95 3.0 2907 2 A57278 fibrillin-2 precur  
40 95 3.0 3871 2 T22812 hypothetical prote  
41 94.5 3.0 956 1 A46016 thrombospondin 3 -  
42 94 2.9 381 1 OOFF2 opsin 2 - fruit fl  
43 94 2.9 612 2 A54282 reversed polarity  
44 94 2.9 915 2 A49874 metabotropic gluta  
45 93.5 2.9 385 2 S53718 homeotic protein d

#### ALIGNMENTS

##### RESULT 1

T27628

hypothetical protein ZC506.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T27628

R:Harris, B.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z20395

A:Accession: T27628

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-999 <WIL>

A:Cross-references: EMBL:Z47073; PIDN:CAA87374.1; GSPDB:GN00028; CESP:ZC506.4

A:Experimental source: clone ZC506

C:Genetics:

A:Gene: CESP:ZC506.4

A:Map position: X

A:Introns: 32/2; 65/3; 77/3; 113/2; 140/3; 177/2; 240/2; 277/1; 315/2; 439/2; 471/3;

C:Superfamily: metabotropic glutamate receptor 4

Query Match 4.3%; Score 136; DB 2; Length 999;

Best Local Similarity 22.3%; Pred. No. 0.035;

Matches 48; Conservative 44; Mismatches 87; Indels 36; Gaps 8;

Qy 354 CKAGFYHPCVLPVNNFRRGDPDQHSSTKDVSEAYV-----CLPCREGCPFCADDSPC 408

Db 615 CKIGFR-----KOLIKDEQCWCACSKCEDVEYLINETHCVGCEQGMWPTDKRGK 664

Qy 409 F---VOEDKYLRL-----AIIISFOGLCMLLDFVSMVYVYHFKAKSIRASGLILLET 457

Db 665 FDLISUSQKYMWRWSMYSLVPTILAVFGIATFLVIVVVIYN--ETPVKASGRELVI 722

Qy 458 ILFSLILLYFPVILFPFSTFRICLLRWALLGFATVGTVTLKLRVLFVLSRTAQR 517

Db 723 LLISIMVICYMTFVLLSRPSAIVCAIKRTGTGFAFSLYSAMFVKTNRIFRIFSTRSAQR 782

Qy 518 ---IPYMTGGVRMRMLAVILLY--VFWEFLI---GW 544

Db 783 PRFISPIQVQVNTAMLAGVQLIGLSLWLSVVPWG 817

##### RESULT 2

I59362

calcium/polyvalent cation-sensing receptor precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999

C:Accession: I59362; A55594

R:Ruut, M.; Molliver, M.E.; Snowman, A.M.; Snyder, S.H.

Proc. Natl. Acad. Sci. U.S.A. 92, 3161-3165, 1995

A:Title: Calcium sensing receptor: molecular cloning in rat and localization to nerve

A:Reference number: I59362; MUID:95241465; PMID:7724534





QY 436 LVYHPRKASIRASGLILLETILFGLSLLLYFPVILYFEPSTFCILLRWARLLGFATV 495  
 Db 642 GVFIKRNPIVKNATNRELSYLLSLCCFSSSLFFIGEPDWTCLRLQPAFGISFVLC 701  
 QY 496 YGVTVTLKLRVLKFLSRTPAQRPYMTGGRVRLAV-----ILLVVFELIGHTSSV 548  
 Db 702 ISCILVKTNRVLVLFVFAKIPTEFHRKRWGNLQFLVFLCTFMQIVICWILYTAPSSY 761  
 QY 549 -CNLEKQISLI-----GQKTSDFHIFNCLI 575  
 Db 762 RNOELEDLFIITCHGSLMALGFLIGYTCLL 793  
 RESULT 7  
 JH0561  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 13-Sep-1998  
 C:Accession: JH0561  
 R:Hanabe, Y.; Masu, M.; Ishii, T.; Shigemoto, R.; Nakanishi, S.  
 Neuron 8, 169-179, 1992  
 A:Title: A family of metabotropic glutamate receptors.  
 A:Reference number: JH0561; MUID:92110002; PMID:1309649  
 A:Accession: JH0561  
 A:Molecule type: mRNA  
 A:Residues: 1-872 <TAN>  
 A:Experimental source: brain  
 C:Comment: This protein is coupled to a G protein and evokes a variety of functions by m  
 C:Superfamily: metabotropic glutamate receptor 4  
 C:Keywords: G protein-coupled receptor; glycoprotein; phosphoprotein; transmembrane prot  
 F:1-18/Domain: signal sequence #status predicted <SIG>  
 F:19-872/Product: metabotropic glutamate receptor 2 #status predicted <MET>  
 F:568-590/Domain: transmembrane #status predicted <TR>  
 F:605-625/Domain: transmembrane #status predicted <TR>  
 F:637-655/Domain: transmembrane #status predicted <TR>  
 F:680-700/Domain: transmembrane #status predicted <TR>  
 F:726-747/Domain: transmembrane #status predicted <TR>  
 F:761-782/Domain: transmembrane #status predicted <TR>  
 F:795-819/Domain: transmembrane #status predicted <TR>  
 F:803-826,338,402,547/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:601,675,827,837,843/Binding site: phosphate (Ser) (covalent) #status predicted  
 F:832/Binding site: phosphate (Thr) (covalent) #status predicted  
 Query Match 3.6%; Score 116; DB 2; Length 872;  
 Best Local Similarity 18.9%; Pred. No. 1;  
 Matches 93; Conservative 63; Mismatches 184; Indels 152; Gaps 22;  
 141 QNKRSR-----QNLQDLDTQALVWSLLE-----GEPSISRAAITFSDLSAPAPQVF 191  
 Db 271 RSEDARELLAATQRLNASTFTWVSDGWGALESVAVASERAEAGITIELASYPIISDFASY 330  
 QY 192 LQATREESRILLQDLSSAPHLANATLETWFPHGLRRKWRPHLRRGNPQGRGLGHSWR 251  
 Db 331 FGS-----LDPWNS-----RNPWF-----REFEERFH-----CSFR 358  
 QY 252 RDKGLGDKSHFKPPYLECENGSKPKGLWLTLSAIYGLQPNLVP----- 298  
 Db 359 ORD-----CAASHLRAVPPEQESKINEVNAVYMAHALHNMHRALCPNTHLCDAMRPVN 414  
 QY 299 -----EFGVMKVINDINQKVDI-DQCSDDGWFSCTHKLNNSECMPIKGLGFLGAYE 351  
 Db 415 GRRLYKDFVNLVNFAPFRPADTDDEVFRDFRGD-----IGRYN 454  
 QY 352 CIC-----KAGFYHFGVLPVNF-----RRRGPDQHSIGSTKDYSE 387  
 Db 455 IFTYLRASGRYRYOKYVWAGBLTDLTFSIPWSPSAGPLPASRCSEPCQLQNEVKRSQVP 514  
 QY 388 E---AYVCLPCREGCPF-----CADD-----SPCFQEDKYLR----- 417  
 Db 515 GEVCCWLCIPQ---PYEVLDEFTCADGLGYWPNASLTGCFELPQYIRWGDARWAVCP 571  
 QY 418 LAISFQGLMCLLDVMSLVYVHFRKASIRASGLILLETILFGLSLLLYFPVILYFEP 477

Db 572 VTACIGALATL--FVLGVFVRH-NATPVVKASGRELCYLLGGVFLCYCMTFVFTAKPS 628  
 QY 478 TERCILLRWARLLGFATVGTVTLKLRVLKVF--LSRTAQRIPIYMT-GGRVRLAVI- 533  
 Db 629 TAVCTLRRLRGLGTAFSCVYSALLTNTNRITRIFGGAREGAQRPRFISPASQVAICLALIS 688  
 QY 534 ---LLVVFWEFLI 542  
 Db 689 GQLLIIVAALVV 700  
 RESULT 8  
 A41939  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
 C:Accession: A41939; S15362  
 R:Houamed, K.M.; Kuiper, J.L.; Gilbert, T.L.; Haldeman, B.A.; O'Hara, P.J.; Mulvihill  
 Science 252, 1318-1321, 1991  
 A:Title: Cloning, expression, and gene structure of a G protein-coupled glutamate rec  
 A:Reference number: A41939; MUID:92022526; PMID:1656524  
 A:Accession: A41939  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1199 <HOU>  
 A:Cross-references: GB:M61099; NID:g397806; PIDN:AAA19497.1; PID:g204460  
 A:Experimental source: cerebellum  
 A:Note: sequence extracted from NCBI backbone (NCBIP:60785)  
 R:Masu, M.; Tanabe, Y.; Tsuchida, K.; Shigemoto, R.; Nakanishi, S.  
 Nature 349, 760-765, 1991  
 A:Title: Sequence and expression of a metabotropic glutamate receptor.  
 A:Reference number: S15362; MUID:91156047; PMID:1847995  
 A:Accession: S15362  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1199 <MAS>  
 A:Cross-references: EMBL:X57569; NID:g56646; PIDN:CAA40799.1; PID:g56647  
 C:Keywords: G protein-coupled receptor; transmembrane protein  
 Query Match 3.6%; Score 115; DB 2; Length 1199;  
 Best Local Similarity 20.7%; Pred. No. 1.8; Mismatches 108; Indels 80; Gaps 9;  
 Matches 61; Conservative 45;  
 357 GFYHPGVLPVNNFRRRQPDQHSIGSTKDYSE-----AYVCLPCREG- 398  
 Db 498 GTHWEGVLNIDYKIQ---MKNKSGMVRSCVSEPCLGKQIKVIRKGEVSCWICTACKENE 554  
 QY 399 -----CPFC-----ADSPCFVQEDKYLR-----LATISFQGLCMLLDFVSMV 438  
 Db 555 FVQDEFTCRACDLGWPNNAELTGCEPIPVRYLEWSDIESIIAIAFSCILGILVTLFVTLIF 614  
 QY 439 YHFRKASIRASGLILLETILFGLSLLLYFPVILYFEPSTFCILLRWARLLGFATVYGT 498  
 Db 615 VLYRDPVYKSSRELXIIILAGIFLVYVCPPTLTAQPTTTTCYLRLLVGLSSAMCYSA 674  
 QY 459 VTLKLRVLKVF-----FLSRTAQRIPIYMTGGRVRLAVIILVVFWEFLIGW 544  
 Db 675 LVTKNTRITRILAGSKKIKCRKPRFMSAWAQVITASILISVOLTLVTLII----- 726  
 QY 545 TSSVCONLEKQISLIGQKTSID-HLIFNNCLIDRWYMTAVGMSLSVSDGLTI 597  
 Db 727 -----MEPPMPILSYPSIKEYILCN-----TSNLGVVAVPVGVNGLLI 764  
 RESULT 9  
 JC2132  
 C:Species: Homo sapiens (man)  
 C:Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 05-Jan-1996  
 C:Accession: JC2132  
 R:Minakami, R.; Katsuki, F.; Yamamoto, T.; Nakamura, K.; Sugiyama, H.  
 Biochem. Biophys. Res. Commun. 199, 1136-1143, 1994  
 A:Title: Molecular cloning and the functional expression of two isoforms of human met

A:Reference number: JC2131; MUID:94197696; PMID:7908515

A:Accession: JC2132

A:Molecule type: mRNA

A:Residues: 1-1180 <MTN>

C:Comment: This protein is coupled to guanine nucleotide binding proteins.

C:Keywords: glycoprotein; neurotransmitter; receptor; transmembrane protein

F:580-604/Domain: transmembrane #status predicted <TM1>

F:617-637/Domain: transmembrane #status predicted <TM2>

F:644-664/Domain: transmembrane #status predicted <TM3>

F:694-714/Domain: transmembrane #status predicted <TM4>

F:738-759/Domain: transmembrane #status predicted <TM5>

F:773-794/Domain: transmembrane #status predicted <TM6>

F:803-827/Domain: transmembrane #status predicted <TM7>

Query Match 3.6%; Score 114; DB 2; Length 1180;

Best Local Similarity 20.4%; Pred. No. 2.1;

Matches 80; Conservative 53; Mismatches 165; Indels 94; Gaps 16;

Qy 219 ETEWPHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGDKSHFKWSPPLYECE 273

Db 333 DVKWFDDYYLKLRPETNHRNP-----WFQEFWQHRFQCRLEGFPOENSKYKNT-----CN 382

Qy 274 NG-----SYKPGWLVTLSAIYGL---QPNLVPEFRGV-----MKVDINLQ 311

Db 383 SSLTLKTHVQDSKMGFVINAIYSMAYLHNNQMSLCPGYAGLCDAMKPIDGRKLLSLM 442

Qy 312 KVIDQCSGDGWFSTHCKHLNNSCMPITKGLGVFLGAYECICKAGFYHGPVLPVNNFR 371

Db 443 KTNFTGVSGDTILFDENGSDSPRYEIMNFKMG-----KDYFDYINVGSDNGEL 492

Qy 372 RGPDOHI-----SGSTKDVSEE-----AYVCLPCREG-----CPFC 402

Db 493 KMDDEVMSKSNIIIRVSCSEPCKEGQIKVIRKGBVSCCWTCTPCKENEYVDEYTKAC 552

Qy 403 -----ADD-SPCFVQEDKYL-----LAIISFOGLCMLLDFVSMVYVYHFRKAKSIRA 449

Db 553 QLGSWPTDGLTCDLIPVOYLRWGDEPIAAVAVFACLGALLATLFTVTVFIIYRDPVVK 612

Qy 450 SGLILETILFGLSLLLYFPVVIYEPSTFCILLRWALLGFATVYGTVTLLKHLRVLV 509

Db 613 SSRELCYIILAGICIGLYCTFCLIAKPKQIYCYLQRIIGLSPAMSYALVTKTNRIARI 672

Qy 510 FLSTRTAQRI-----PYMTGGRVNRMLAVILLV 537

Db 673 -LAGSKKKICTKKPRFMSACQALVIAFILICI 703

RESULT 10

JC2131

Species: Homo sapiens (man)

C:Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 05-Jan-1996

C:Accession: JC2131

R:Minakami, R.; Katsuki, F.; Yamamoto, T.; Nakamura, K.; Sugiyama, H.

Biochem. Biophys. Res. Commun. 199, 1136-1143, 1994

A:Title: Molecular cloning and the functional expression of two isoforms of human metabo

A:Reference number: JC2131; MUID:94197696; PMID:7908515

A:Accession: JC2131

A:Molecule type: mRNA

A:Residues: 1-1212 <MTN>

C:Comment: This protein is coupled to guanine nucleotide binding proteins.

C:Keywords: glycoprotein; neurotransmitter; receptor; transmembrane protein

F:580-604/Domain: transmembrane #status predicted <TM1>

F:617-637/Domain: transmembrane #status predicted <TM2>

F:644-664/Domain: transmembrane #status predicted <TM3>

F:694-714/Domain: transmembrane #status predicted <TM4>

F:738-759/Domain: transmembrane #status predicted <TM5>

F:773-794/Domain: transmembrane #status predicted <TM6>

F:803-827/Domain: transmembrane #status predicted <TM7>

Query Match 3.6%; Score 114; DB 2; Length 1212;

Best Local Similarity 20.4%; Pred. No. 2.2;

Matches 80; Conservative 53; Mismatches 165; Indels 94; Gaps 16;

Qy 219 ETEWPHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGDKSHFKWSPPLYECE 273

Db 333 DVKWFDDYYLKLRPETNHRNP-----WFQEFWQHRFQCRLEGFPOENSKYKNT-----CN 382

Qy 274 NG-----SYKPGWLVTLSAIYGL---QPNLVPEFRGV-----MKVDINLQ 311

Db 383 SSLTLKTHVQDSKMGFVINAIYSMAYLHNNQMSLCPGYAGLCDAMKPIDGRKLLSLM 442

Qy 312 KVIDQCSGDGWFSTHCKHLNNSCMPITKGLGVFLGAYECICKAGFYHGPVLPVNNFR 371

Db 443 KTNFTGVSGDTILFDENGSDSPRYEIMNFKMG-----KDYFDYINVGSDNGEL 492

Qy 372 RGPDOHI-----SGSTKDVSEE-----AYVCLPCREG-----CPFC 402

Db 493 KMDDEVMSKSNIIIRVSCSEPCKEGQIKVIRKGBVSCCWTCTPCKENEYVDEYTKAC 552

Qy 403 -----ADD-SPCFVQEDKYL-----LAIISFOGLCMLLDFVSMVYVYHFRKAKSIRA 449

Db 553 QLGSWPTDGLTCDLIPVOYLRWGDEPIAAVAVFACLGALLATLFTVTVFIIYRDPVVK 612

Qy 450 SGLILETILFGLSLLLYFPVVIYEPSTFCILLRWALLGFATVYGTVTLLKHLRVLV 509

Db 613 SSRELCYIILAGICIGLYCTFCLIAKPKQIYCYLQRIIGLSPAMSYALVTKTNRIARI 672

Qy 510 FLSTRTAQRI-----PYMTGGRVNRMLAVILLV 537

Db 673 -LAGSKKKICTKKPRFMSACQALVIAFILICI 703

RESULT 11

JH0562

metabotropic glutamate receptor 3 precursor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 13-Sep-1998

C:Accession: JH0562

R:Tanabe, Y.; Masu, M.; Ishii, T.; Shigemoto, R.; Nakanishi, S.

Neuron 8, 169-179, 1992

A:Title: A family of metabotropic glutamate receptors.

A:Reference number: JH0561; MUID:92110002; PMID:1309649

A:Accession: JH0562

A:Molecule type: mRNA

A:Residues: 1-879 <TAN>

A:Experimental source: brain

C:Comment: This protein is coupled to a G protein and evokes a variety of functions b

C:Superfamily: metabotropic glutamate receptor 4

C:Keywords: G protein-coupled receptor; glycoprotein; phosphoprotein; transmembrane p

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-879/Product: metabotropic glutamate receptor 3 #status predicted <MET>

F:577-599/Domain: transmembrane #status predicted <TRI>

F:614-634/Domain: transmembrane #status predicted <TII>

F:646-664/Domain: transmembrane #status predicted <III>

F:689-709/Domain: transmembrane #status predicted <TIV>

F:735-756/Domain: transmembrane #status predicted <TVI>

F:770-791/Domain: transmembrane #status predicted <TVI>

F:804-828/Domain: transmembrane #status predicted <VII>

F:209, 292, 414, 439/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:610, 845/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 3.5%; Score 112; DB 2; Length 879;

Best Local Similarity 22.6%; Pred. No. 2.1;

Matches 42; Conservative 30; Mismatches 76; Indels 38; Gaps 7;

Qy 390 YVCLPCRE-----GCPCF-----ADSPCFVQEDKYL-----LAISF 423

Db 529 WICIPCEPYELVDFTCMDCGPGQWPTADLSGCYNLPEDYIKWEDANAIGPVTIACLG 588

Qy 424 QGLCMLDFVSMVYVYHFRKAKSIRASGLILETILFGLSLLLYFPVVIYEPSTFCIL 483

Db 589 LCTCIVI-----TVFIKHNNTPLVKASGRELCYILFGLSVLSYCWTFFFIAKPSVICAL 643

Qy 484 LRWARLLGFATVYGTVTLLKHLRVLVKVF---LSRTAQRIPIYMT---GGRVMRLAVIL---LV 536

Db 644 RRLGLTGFSAICYALLTKTNCIARIIFDGVKGAQRKPFISPSQVFCILGLILVQIVMW 703  
 QY 537 VFWFLI 542  
 Db 704 SVWLIL 709  
 RESULT 12  
 A42916  
 metabotropic glutamate receptor mGluR5 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
 C:Accession: A42916  
 R:Abel, T.; Sugihara, H.; Nawa, H.; Shigemoto, R.; Mizuno, N.; Nakanishi, S.  
 J. Biol. Chem. 267, 13361-13368, 1992  
 A:Title: Molecular characterization of a novel metabotropic glutamate receptor mGluR5  
 A:Reference number: A42916; MUID:92317054; PMID:1320017  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1171 <ABE>  
 A:Cross-references: GB:dl0891; NID:g220813; PIDN:BAA01711.1; PID:dl002186; PID:g220814  
 A:Experimental source: brain  
 A:Note: sequence extracted from NCBI backbone (NCBI:107749, NCBI:107750)  
 C:Keywords: G protein-coupled receptor; transmembrane protein  
 Query Match 3.5%; Score 111; DB 2; Length 1171;  
 Best Local Similarity 20.8%; Pred. No. 3.6;  
 Matches 80; Conservative 54; Mismatches 171; Indels 80; Gaps 17;  
 QY 219 ETEWFGHURKWRPHLHRRGNQGRGLGHSWR-----RKDGLGDKSHFKWSPPYLECE 273  
 Db 332 DVKWFDDYLLKRPETNLNRP-----WFOEFWQHRFQCRLECFQAENSKYNT-----CN 381  
 QY 274 NG-----SYRPGWLV-TLSSAIYGL---QPNLVPEFRGV-----MKVDINLQ 311  
 Db 382 SSSLTRTHVQDSKMGFVINAIYSMAVGLHNMQMSLCPGYAGLCDAMKPIDGRKLLDSL 441  
 QY 312 KVIDQCSDDGWFSTGTHKCHLNSECMPIKGLGFLVGAECICKAGFVHCVLPV----- 366  
 Db 442 KTNFTVSGDMILFDENGDCPRVEIMNFKEG---KDYFDYINVGSWDNGLKMDDEEV 498  
 QY 367 ----NFRRRGPDQHI-SGSTKDVSEE----AYVCLPREG-----CPFC-----A 403  
 Db 499 WSKNNIIRSVCEPKCKGQIKVIRKGVSCCWTCTPKCKENYVFEVDEYTCACQLGSWPT 558  
 QY 404 DD-SPCFVQEDKYLK-----LAISFGGLCMLLDFVSMVYHYHFRKAKSTRASGLLILLE 456  
 Db 559 DDLTGCDLIPVQYLRWGDPEPIAAWVFACGLGLATLFTVTFIITYRTPVVKSSSRELCY 618  
 QY 457 TILFGSLLLYPPVVLVPEPSTFCILLRWARLLGFATVYGTVTLKHLRVLKVFLSRTAQ 516  
 Db 619 IILAGICLGLYCTCLLAKPKQIYCYLQRIQIGLSPMSYSALVTKTNRIARI-LAGSKK 677  
 QY 517 RI----PYMTGGRVRLMAVILLVV 537  
 Db 678 KICTKKPRMSACAQLVIAFLICI 702  
 RESULT 13  
 T34513  
 hypothetical protein ZK783.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T34513  
 R:Favell, A.; Vaudin, M.  
 submitted to the EMBL Data Library, August 1994  
 A:Description: The sequence of C. elegans cosmid ZK783.  
 A:Reference number: Z21536  
 A:Accession: T34513  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-3507 <FAV>

A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1  
 A:Experimental source: strain Bristol N2; clone ZK783  
 C:Genetics:  
 A:Gene: CESP:ZK783.1  
 A:Map position: 3  
 A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/3504/1  
 Query Match 3.5%; Score 110.5; DB 2; Length 3507;  
 Best Local Similarity 18.0%; Pred. No. 16;  
 Matches 76; Conservative 51; Mismatches 137; Indels 159; Gaps 19;  
 QY 20 GAVGASRDPOGRPDSP-----RERTPKG-----KPHAQOQFRA 52  
 Db 1227 GDNGETSGVDGKPTTAPTSSAESSTSRIPPTSEASPEGSGEAGVPESPDGSSEST 1286  
 QY 53 SASDSSAPWSRSTDTGTLAOKLAEEVPMVAVSYLYTGDHOLKRNCSGRYELAGLPGKW 112  
 Db 1287 SAPDGVSTSSAT-----APEVPTTSAS--STPDA-----VEESGIPS-- 1322  
 QY 113 PALASAPSLHRLDITLTHATNFMVLOSNNKSRQNLQDDLDWYQALVWSLLEGEPSIS 172  
 Db 1323 TSKTAP-----LETTAPSTEVTSPT--EGSGTEESTLPTT-----EGSGESTTSAPTIVE 1371  
 QY 173 RAAITFTDLSAPAPQVFLQATREESRILLQDLSSAPHLANATLETWFHGLRRKWRP 232  
 Db 1372 PATVLPQNRNEKPEP-----TKDTFALPTTTTGAPOANDSSVE----- 1410  
 QY 233 HLHRRGNQGRGLGHSWRKDKGLGDKSHFKWSPPYLECENGSYKPGWLVTLSAIYGL 292  
 Db 1411 --NTKCTSSDCGLDALCERRTGVCRCPEGFGAPPKKSC----- 1448  
 QY 293 QPNLVPEFRGVKMDINLQKVIDQCSDDGWFSTGTHKCHLNSECMPIKGLGFLVGAEC 352  
 Db 1449 -----VDVDECA-----TGDHNC--ESARCON-----YVGGYAC 1476  
 QY 353 ICKAGFYHGVLPVNNFRRRGPDOHISGSTKDVSEAYVCLPREGPFCADDSPCFVQE 412  
 Db 1477 FCPTGF-----RKADD-----GSCQDIDE-----CTEHNSTC--CGANAKCVNKP 1514  
 QY 413 DKY 415  
 Db 1515 GTY 1517  
 RESULT 14  
 JC5701  
 Erbb kinase activator alpha, brain and thymus - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 25-Nov-1997 #sequence\_revision 25-Nov-1997 #text\_change 02-Aug-2002  
 C:Accession: JC5701; PC4411  
 R:Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; M1  
 J. Biochem. 122, 675-680, 1997  
 A:Title: A novel brain-derived member of the epidermal growth factor family that inte  
 ating the differentiation of MDA-MB-453 cells.  
 C:Superfamily: human Erbb Kinase activator alpha, brain and thymus; EGF homology  
 F;361-397/Domain: EGF homology <EGF>  
 Query Match 3.4%; Score 109.5; DB 2; Length 868;  
 Best Local Similarity 19.1%; Pred. No. 3.2;  
 Matches 92; Conservative 57; Mismatches 173; Indels 159; Gaps 20;  
 QY 31 RPDSPRERTPKGPHAQOP-----GRASAS-----DSSAPWSRSTDTGTLA----- 71

```
Db 74 RPAAPPEPRPOPOPRKSPARAAARAAAGGMRDPAPGSSMLLFGVSLACYSPSL 133
QY 72 --QKLAEEVPMDSVLY-----TGDHQLKRNCSGRYELAGLPKGPALASAHPSL 122
Db 134 KSVODQAYKAPVVVEGKVOGLAPAGSSSNSTREPPASGRVALVKVLDKWP----- 184
QY 123 HRALDTLTHATNLFNLVMSKREONLQ-----DLDLWQALVMSLLEGEPSISRAAITF 178
Db 185 -----LRSGGLREQVISVGCAPLERNQRYIFEL---EP--TEQPLVF 223
QY 179 STDLSAPAPQVFLQATREESRILLDLSSAPHLANATLETFWHLRKRKWRPHLRRC 238
Db 224 KT--AFAPVDNPKNIKKEVGKILTCDCAT-----RPKLKKMK 259
QY 239 PNQGRGLGHSWRKDKGLGDKSHFKWSPPYLECENGSKYKPGWLVTLSSAI---YG--- 291
Db 260 SQTGEVGEKOSLCEAAGNPQPSYRW-----FKDGKELNRSRDIRIKYNGR 308
QY 292 ---LQPNLVP-----EFGVMKVDINLQKVIDOCSS---DGMFSGTHKCH----- 331
Db 309 NSRLQFNKVKVEDAGEVCEAEINILCKDTVRGLHVNVSVTLLSSWSGHARKCNETAKSY 368
QY 332 -LNNSECMPIKGLGFLVAYECICKAGFYHPGVLPVNNFRRGPD---QHSISTKDVSE 387
Db 369 CVNGGVCIYIEG-----INLSCKCPNGFQGRCLERLPLRLYMPDKQKHLGFEKRE-AE 423
QY 388 EAYVCLPCREGCPFCADDSPCFVQEDKYLRLATISFGCLMLLDVYMLVYVYHFRKAKSI 447
Db 424 ELY-----QKRVLTITGICVALLVGVICVVAICTKKQ 457
QY 448 R 448
Db 458 R 458

RESULT 15
S53611
MIBP1 protein - rat
N:Alternate names: angiotensinogen gene-inducible enhancer-binding protein; c-myc intron
C:Species: Rattus norvegicus (Norway rat)
C>Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 20-Jun-2000
C:Accession: S53611; A39796; S22292; I58280
R:Makino, R.; Akiyama, K.; Yasuda, J.; Mashiyama, S.; Honda, S.; Sekiya, T.; Hayashi, K.
Nucleic Acids Res. 22, 5679-5685, 1994
A:Title: Cloning and characterization of a c-myc intron binding protein (MIBP1).
A:Reference number: S53611; MUID:95140632; PMID:7838722
A:Accession: S53611
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2437 <MAK>
R:Cross-references: EMBL:D37951; NID:g1408559; PIDN:BAA07168.1; PID:g662296
R:Ron, D.; Brasier, A.R.; Habener, J.F.
Mol. Cell. Biol. 11, 2887-2895, 1991
A:Title: Angiotensinogen gene-inducible enhancer-binding protein 1, a member of a new fa
A:Reference number: A39796; MUID:91203912; PMID:2017183
A:Accession: A39796
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1,1523-2263,'K',2265-2437 <RON>
R:Cross-references: GB:M65251; NID:g202790; PIDN:AAA40698.1; PID:g202791
R:Mitchemore, C.; Traboni, C.; Cortese, R.
Nucleic Acids Res. 19, 141-147, 1991
A:Title: Isolation of two cDNAs encoding zinc finger proteins which bind to the alpha 1-
A:Reference number: I58280; MUID:91187610; PMID:1901405
A:Accession: S22292
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 'E',1729,'R',1734-2437 <MIT>
R:Cross-references: EMBL:X54249; NID:g57517; PIDN:CAA38150.1; PID:g57518
A:Note: the authors did not translate the codon for residue 1
C:Superfamily: HIV-Ep2 enhancer-binding protein
C:Keywords: DNA binding; duplication; metal binding; transcription regulation; zinc fing
```

```
F:191-211/Region: zinc finger CCHH motif
F:219-241/Region: zinc finger CCHH motif
F:934-940/Region: nuclear location signal
F:947-979/Region: serine-rich
F:1792-1812/Region: zinc finger CCHH motif
F:1820-1842/Region: zinc finger CCHH motif
F:1890-1917/Region: acidic

Query Match 3.4%; Score 107.5; DB 2; Length 2437;
Best Local Similarity 21.1%; Pred No. 17;
Matches 71; Conservative 54; Mismatches 133; Indels 79; Gaps 15;

QY 18 GLGAVGASRDPOGRDPSRPTPKGPKHAQPGRASDSSAPWSRSTGDTILAQKLAEE 77
Db 1278 GASGLHSHKLNLPKFPSPDGSKSTEAPPTQELLREDFASENAGP-LQSLPCTVVPVRIQTH 1336
QY 78 VPMDSVSYLYTGDHQLKRNCSGRYELAGLPKGPALASAHPSLRALDTLTHATNLFN 137
Db 1337 VP-SYGSVMYTSISQILGONS-----PAIV-----ICKVDENMTQRTLVN 1376
QY 138 VMLOS---NKSREQNLQDDLDLWQALVMSLLEGE-----SISRAAITFEST 180
Db 1377 AMQGGFNIAQVLGORTGLEKYP--LWKVPQTLPLGLESSIPCLPSTSDSAASLGGSK 1434
QY 181 DLSAPAPQVFLQATREESRI-----LLQDLSSAPHLANATLETFWHLRKRKWRP 232
Db 1435 RMLSPASSLELFMETKQKRVKEEKMYGOIVEEL--SAVELNDSIK-----KGLSRQKP 1488
QY 233 HLHRRPNQGRGLGHSWRKDKGL-----GGDKSHFKWS---PPYLECENGSKYKPGWLV 284
Db 1489 QLVROGCASEPKD-GSSQSRSSFSLSLSPSSQDHPAASGPFPPNREILSGSRAP----- 1542
QY 285 LSSAIYGLQPNLVPEPRGVKVDINLQKVIDOCSSD 321
Db 1543 -----PR--RKFSGPSSESSELDIDETSSD 1568
```

Search completed: February 15, 2003, 08:32:14  
Job time : 27.1803 secs





GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2003, 04:22:33 ; Search time 9.24587 Seconds  
(without alignments)  
2687.074 Million cell updates/sec

Title: US-09-775-181-4  
Perfect score: 3199  
Sequence: 1 MCMAYPLLCLLLAQLGLG.....YMTAVGMWSLVSYDGLTIFQ 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	161.5	5.0	MGR_DROME	P91685 drosophila
2	136	4.3	MGR1_CAEEL	Q09630 caenorhabdi
3	134.5	4.2	CASR_MOUSE	Q9G996 mus musculus
4	133.5	4.2	CASR_RAT	P48442 rattus norv
5	127	4.0	CASR_HUMAN	P41180 homo sapien
6	123	3.8	CASR_BOVIN	P35384 bos taurus
7	120	3.8	MGR2_HUMAN	Q14416 homo sapien
8	116	3.6	MGR2_RAT	P31421 rattus norv
9	115	3.6	MGR1_HUMAN	Q13255 homo sapien
10	115.5	3.6	MGR3_HUMAN	Q14832 homo sapien
11	115	3.6	MGR1_RAT	P23385 rattus norv
12	114	3.6	MGR5_HUMAN	P41594 homo sapien
13	112	3.5	MGR3_RAT	P31422 rattus norv
14	111	3.5	MGR5_RAT	P31424 rattus norv
15	109.5	3.4	NRG2_RAT	O35569 rattus norv
16	106.5	3.3	NRG2_HUMAN	Q14511 homo sapien
17	106	3.3	DLK_HUMAN	P80370 homo sapien
18	104.5	3.3	UROM_RAT	P27590 rattus norv
19	104.5	3.3	TFP3_HUMAN	P49746 homo sapien
20	104.5	3.3	CLR3_HUMAN	Q9NYQ7 homo sapien
21	102.5	3.2	SGA5_HUMAN	Q9Y345 homo sapien
22	101	3.2	MGR4_HUMAN	Q14833 homo sapien
23	100.5	3.1	CLR3_RAT	O89278 rattus norv
24	100	3.1	DCUP_DROME	Q9V595 drosophila
25	99.5	3.1	NTC2_HUMAN	Q04721 homo sapien
26	99	3.1	MM15_HUMAN	P51511 homo sapien
27	98.5	3.1	VFN1_PIG	P48819 sus scrofa
28	98	3.1	FBN1_BOVIN	P98133 bos taurus
29	97.5	3.0	COA2_PAVEN	P22964 porcine par
30	97.5	3.0	FBN1_HUMAN	P35555 homo sapien
31	97	3.0	MGR7_HUMAN	Q14831 homo sapien
32	97	3.0	S24A_HUMAN	O95486 homo sapien
33	97	3.0	FBN2_HUMAN	P35556 homo sapien

RESULT 1				
ID	MGR_DROME	STANDARD	PRT	976 AA
AC	P91685: Q9V485;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DE	Metabotropic glutamate receptor precursor.			
GN	GLU-RA OR GLURA OR CG11144.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;			
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Oregon-R;			
RX	MEDLINE=96421661; PubMed=8824309;			
RA	Parmentier M.L., Pin J.P., Bockaert J., Grau Y.;			
RT	"Cloning and functional expression of a Drosophila metabotropic			
RT	glutamate receptor expressed in the embryonic CNS.";			
RL	J. Neurosci. 15:6687-6694(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Berkley;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abriel J.F., Agbayani A., An H.-J., Andrews-pfannkoch C., Baldwin D.,			
RA	Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,			
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Delcher A., Dahlke C., Davenport L.B., Davies P.,			
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,			
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,			
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,			
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,			
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,			

## ALIGNMENTS



```
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT TRANSMEM 682 704 POTENTIAL.
FT TRANSMEM 719 739 POTENTIAL.
FT TRANSMEM 751 769 POTENTIAL.
FT TRANSMEM 792 812 POTENTIAL.
FT TRANSMEM 836 857 POTENTIAL.
FT TRANSMEM 871 893 POTENTIAL.
FT TRANSMEM 904 929 POTENTIAL.
FT DOMAIN 888 891 POLY-PHE.
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 999 AA; 113275 MW; AB22AF2A28D9A0CF CRC64;

Query Watch 4.3%; Score 136; DB 1; Length 999;
Best Local Similarity 22.3%; Pred. No. 0.013;
Matches 48; Conservative 44; Mismatches 87; Indels 36; Gaps 8;

QY 354 CKAGFYHGVLPVNNFRRRGPDQHSSTKDVSEAYV-----CLPCRGCGPCADDSPC 408
DB 615 CKIGFR-----KOLIKDQCQWACKSCDEYELINEHVCVCEQGWPTKDKGC 664
QY 409 F---VOEDKYLRL-----AIISFGGLCMMLDFVSMVLVYHFRKAKSIRASGLILLET 457
DB 665 FDLISQLKYMWRMSYSLVPTILAVFGIATLFVIVVVIYN--ETPVVKASGRELSYI 722
QY 458 ILFGSLLLPVVILFESTFRCILLRWLRLLGATVTVGVTLKHLRVLKVFLSTAQOR 517
DB 723 LLISIMCMYCTPVLISKPSAIVCAIKRTGIGFAFSLYSAMFVKTNRIFRIFSTRSAQR 782
QY 518 ---IPYMTGGRVMRLAVILLV--VFWEILI---GW 544
DB 783 PRISPISQVWVTAMLAGVOLIGSLIWLVSVPVPGW 817

RESULT 3
CASR_MOUSE STANDARD; PRT: 1079 AA.
AC Q9QY96: O08968; O88519; Q9QY95; Q9QZ08; Q9RID6; Q9RLY2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Extracellular calcium-sensing receptor precursor (CaSR) (Parathyroid
DE Cell calcium-sensing receptor).
GN CASR OR GPRC2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
STRAIN=C57BL/6; TISSUE=Kidney;
MEDLINE=20092890; PubMed=10625662;
Oda Y., Tu C.-L., Chang W., Crumrine D., Koemueves L., Mauro T.,
Elias P.M., Bikle D.D.;
RA "The calcium sensing receptor and its alternatively spliced form in
RT murine epidermal differentiation.";
RL J. Biol. Chem. 275:1183-1190(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.
STRAIN=Black Swiss X 129/SVJ; TISSUE=Kidney;
RX MEDLINE=20119279; PubMed=10652312;
RA Pi M., Garner S.C., Flannery P., Spurney R.F., Quarles L.D.;
RT "Sensing of extracellular cations in CaSR-deficient osteoblasts.
RT Evidence for a novel cation-sensing mechanism.";
RL J. Biol. Chem. 275:3256-3263(2000).
RN [3]
RP SEQUENCE OF 256-600 FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.
TISSUE=Epiphyseal cartilage;
RX MEDLINE=20043955; PubMed=10579354;
Chang W., Tu C., Chen T.-H., Koemueves L., Oda Y., Pratt S.A.,
RA Miller S., Shoback D.;
RT "Expression and signal transduction of calcium-sensing receptors in
RT cartilage and bone.";
RL Endocrinology 140:5883-5893(1999).
```

```

[4]
RN SEQUENCE OF 507-582 FROM N.A. (ISOFORM A).
RP STRAIN=NMRI; TISSUE=Brain;
RA Hildenbrand J., Ammon H.P.T., Wahl M.A.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE OF 562-814 FROM N.A.
RP TISSUE=Kidney;
RA Moawad T.I., Riccardi D.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[6]
RN SEQUENCE OF 646-799 FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=97231187; PubMed=9076582;
RA Quarles L.D., Hartle J.E. II, Siddhanti S.R., Guo R., Hinson T.K.;
RT "A distinct cation-sensing mechanism in MC3T3-E1 osteoblasts
RT functionally related to the calcium receptor.";
J. Bone Miner. Res. 12:393-402(1997)
CC -!- FUNCTION: SENSES CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
CC MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPIDERMIS, KIDNEY AND CARTILAGE.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
-----
EMBL; AF110178; AAD28371.1; -
DR EMBL; AF110179; AAD28372.1; -
DR EMBL; AF128842; AAD40638.1; -
DR EMBL; AF068900; AAC19388.1; -
DR EMBL; AB027140; BAA77688.1; -
DR EMBL; AF002015; AAC53252.1; -
DR EMBL; AF159565; AAF00193.1; -
DR MGI; MGI:1351351; Gprc2a.
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00379; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
DR PROSITE; PS0259; G_PROTEIN_RECEP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1079
FT DOMAIN 20 612
FT TRANSMEM 613 635
FT DOMAIN 636 649
FT TRANSMEM 650 670
FT DOMAIN 671 681
FT TRANSMEM 682 700
FT DOMAIN 701 724
FT TRANSMEM 725 745
FT DOMAIN 746 769
FT TRANSMEM 770 792
FT DOMAIN 793 805
FT TRANSMEM 806 828
FT DOMAIN 829 836
FT TRANSMEM 837 862
FT DOMAIN 863 1079
FT CARBOHYD 90 90
FT CARBOHYD 130 130
POTENTIAL.
EXTRACELLULAR CALCIUM-SENSING RECEPTOR.
EXTRACELLULAR (POTENTIAL).
I (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
II (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
III (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
IV (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
V (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
VI (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
VII (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT CARBOHYD 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 287 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 446 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 541 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 594 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 461 MISSING (IN ISOFORM B).
FT CONFLICT 45 A -> S (IN REF. 2).
FT CONFLICT 304 L -> P (IN REF. 3).
FT CONFLICT 410 G -> D (IN REF. 2 AND 3).
FT CONFLICT 566 V -> A (IN REF. 2).
FT CONFLICT 595 Y -> H (IN REF. 2).
FT CONFLICT 610 E -> V (IN REF. 5).
FT CONFLICT 814 F -> L (IN REF. 5).
FT CONFLICT 889 L -> I (IN REF. 2).
FT CONFLICT 906 TGSN -> SGWI (IN REF. 2).
FT CONFLICT 1057 V -> M (IN REF. 2).
FT CONFLICT 1064 V -> A (IN REF. 2).
FT CONFLICT 1076 I -> V (IN REF. 2).
SQ SEQUENCE 1079 AA; AAF8D8D472736D6E CRC64;

Query Match 4.2%; Score 134.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.019;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGPNQGRGLGSHRRDGLGDKSHKFPSPYLECENGSKPGWLVLSAAYGLQP 294
DB 377 HEEG-----GNRLNLSSTAFRLCTGDENINSVETPYMGYEHRLRISYNYLVAVYSIAHALQD 433
QY 295 --NLVPEFRGVKVDINLQVDIDQSSDGFSGTHKCHLNS-----ECMPIKG 342
DB 434 IYCLPG-RGLF---TNGSCADIKV--FAWQVLKHLRLHNTNNMGEQVTFDECGDLVG 487
QY 343 LGFVL-----GAYECICKAG---FYHGPVLPVNNRRRGPQDHIS----- 379
DB 488 NYSIINHLSPEDSGSVFVEKGVYNYAKKGBLFLNEGKILWSGFSREVPSNCSRDQC 547
QY 380 -GSTKDVSE---AVCLPCREG-----CPFCADD-----SPCFVQEDKYLR 417
DB 548 AGTRKGIIEGPTCCPECCECPDGEYSGETDSACDKPDWSNENYTSIAKEIEFLA 607
QY 418 -----LAIISFGCLMLDFVSMVLVYVFRKAKSIRASGLILLETILFGSLLLYPPVVI 471
DB 608 WTEPFQIALTLFAVLGIFTAVLGVIFKFRTPIVKATNRELSYLLLESICCFSSSLF 567
QY 472 LYFEPSTFCILLRWARLLGFATVYGTVTLKHLRVLVKLSRTAQRIPYMTGCRVNRMLA 531
DB 668 FIEPQDWTCLRLQPAFGISFVLCISCLVKTNRVLLVFEAKIPTSFHRKWWGLNLQFL 727
QY 532 V-----ILLVVFVFLGWTSSVQNLKQISLIGCGKTSDDLHFNMLCLIDRWYMTAV 584
DB 728 VFLCTPMVICIIMLYTAPPSSY-RNHELE-----DEIIFITC---HEGSLMAL 773
QY 585 GMSLVSY 592
DB 774 G--SLIGY 779

RESULT 4
CASR_RAT STANDARD; PRT: 1079 AA.
AC P48442:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Extracellular calcium-sensing receptor precursor (CaSR) (Parathyroid
DE Cell calcium-sensing receptor).
GN CASR OR GPRC2A OR PCARL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
```

```
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Kidney outer medulla;
RA MEDLINE=95116508; PubMed=7816802;
RX Riccardi D., Park J., Lee W., Gamba G., Brown E.M., Hebert S.C.;
RT "Cloning and functional expression of a rat kidney extracellular
calcium/polyvalent cation-sensing receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:131-135(1995).
[2]
RN SEQUENCE OF 1-294 FROM N.A.
RP STRAIN=Wistar;
RX MEDLINE=95241465; PubMed=7724534;
RA Ruat M., Snowman A.M., Snyder S.H.;
RT "Calcium sensing receptor: molecular cloning in rat and localization
to nerve terminals.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3161-3165(1995).
CC -!- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF
CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A
G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND
MESSENGER SYSTEM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
EMBL: U10354; AAC52149.1; -.
DR EMBL; U20289; AAC52195.1; -.
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
DR PROSITE; PS02029; G_PROTEIN_RECEP_F3_4; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1079
FT DOMAIN 20 612
FT TRANSMEM 613 635
FT DOMAIN 636 649
FT TRANSMEM 650 670
FT DOMAIN 671 681
FT TRANSMEM 682 700
FT DOMAIN 701 724
FT TRANSMEM 725 745
FT DOMAIN 746 769
FT TRANSMEM 770 792
FT DOMAIN 793 805
FT TRANSMEM 806 828
FT DOMAIN 829 836
FT TRANSMEM 837 862
FT DOMAIN 863 1079
FT CARBOHYD 90 90
FT CARBOHYD 130 130
FT CARBOHYD 261 261
FT CARBOHYD 287 287
FT CARBOHYD 386 386
FT CARBOHYD 446 446
FT CARBOHYD 468 468
FT CARBOHYD 488 488
FT CARBOHYD 541 541
FT CARBOHYD 594 594
SQ SEQUENCE 1079 AA; 120867 MW; D7664550361F9736 CRC64;
```

```
Query Match 4.2%; Score 133.5; DB 1; Length 1079;
Best Local Similarity 21.7%; Pred. No. 0.023;
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;

QY 235 HRRGNQGRGLGHSRRRDKGLGSKSHFKWSPYLECENGSKPGWLYTLSSAIYGLQP 294
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 377 HEEG---GNRLNLSSTAFRLPCLTGDENINSVETPYMDYEHRLISYNYLVAVSIAHALQD 433
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 295 --NLVPEFRGVKMDINLQKVIDOCSSDGFSGTHKCHLNNS-----ECMPIRG 342
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 434 IYTCLPG-RGLF---INGSCADIKV--EAWQVLKHLRLHNTNNNGEQVTFDECGDLVG 487
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 343 LGFVL-----GAYECICAKG---FYHGVLPVNNFRFRGPDQIHIS----- 379
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 488 NYSIINWHLSPEDGSIVFKEGVGYNYIYAKKGERLFINEEKILMWSGFSEVPFNSCRDQC 547
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 380 -GSTKDVSEB---AVVCLPCEBEG-----CPFCADD-----SPCFVQEDKYL 417
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 548 AGTRKGLIEGPTCCFCEVCEPDGEYSGETDASACDKPDDFWSNENHTSCIAKEIEFLA 607
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 418 -----LAIISFOGLMLDFVSMVYVYHFRKAKSIRASGLILLETILFGSLLLYPVVI 471
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 608 WTEPFGIALTFLFAVLGIFLTAFLGVFIFKFRNTPIVKATNRELSYLLSLFCCFSSSLF 667
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 472 LYEPSTFRCLLRWARLLGFATVGTVLKLRVLKVLFSRTAQRIPIWTGGRVNRMLA 531
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 668 FTGEPQDWTCLRQPAFGISFVLCISCLVKNRVLLVFEAKIPTGFHRKWWGLNLQFL 727
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 532 V-----ILLVFWFLIGTSSVCONLEKQISLIQGKTSOHLIFNMCLIDRWYNTAV 584
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 728 VFLCTFMQLICIIWLYTAPPSSY-RNHELE-----DEIIFITC---HEGSLMAL 773
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 585 GWMSLVSY 592
      | | | |
Db 774 G--SLIGY 779

RESULT 5
CASR_HUMAN STANDARD; PRT: 1078 AA.
AC P41180; Q13912; Q16379; Q16108; Q16109; Q16110;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Extracellular calcium-sensing receptor precursor (CaSR) (Parathyroid
  Cell calcium-sensing receptor).
GN CASR OR GPRC2A OR PCAR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RA Pearce S.H.S., Thakker R.V.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
[2]
SEQUENCE FROM N.A.
RC TISSUE=Parathyroid;
RX MEDLINE=95279439; PubMed=7759551;
RA Garrett J.E., Capuano I.V., Hammerland L.G., Hung B.C., Brown E.M.,
  Hebert S.C., Nemeth E.F., Fuller F.;
RT "Molecular cloning and functional expression of human parathyroid
  calcium receptor cDNAs.";
RL J. Biol. Chem. 270:12919-12925(1995).
[3]
SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=95408281; PubMed=7677761;
RA Aida K., Koishi S., Tawata M., Onaya T.;
RT "Molecular cloning of a putative Ca(2+)-sensing receptor cDNA from
  human kidney.";
RL Biochem. Biophys. Res. Commun. 214:524-529(1995).
[4]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=96343808; PubMed=8756555;
RA Freichel M., Zink-Lorenz A., Holloschi A., Hafner M., Flockerzi V.,
  Raue F.;
RT "Expression of a calcium-sensing receptor in a human medullary
  thyroid carcinoma cell line and its contribution to calcitonin
  secretion.";
RL Endocrinology 137:3842-3848(1996).
[5]
RP SEQUENCE OF 643-908 FROM N.A.
RX MEDLINE=96193893; PubMed=8613532;
RA Bikle D.D., Rainan A., Mauro T., Harris J., Pillai S.;
RT "Changes in calcium responsiveness and handling during keratinocyte
  differentiation. Potential role of the calcium receptor.";
RL J. Clin. Invest. 97:1085-1093(1996).
[6]
RP VARIANTS FHH GLN-185; LYS-297 AND TRP-795.
RX MEDLINE=94094324; PubMed=7916660;
RA Pollak M.R., Brown E.M., Chou Y.-H.W., Hebert S.C., Marx S.J.,
  Steinmann B., Levi T., Seidman C.E., Seidman J.G.;
RT "Mutations in the human Ca(2+)-sensing receptor gene cause familial
  hypocalciuric hypercalcaemia and neonatal severe
  hyperparathyroidism.";
RL Cell 75:1297-1303(1993).
[7]
RP VARIANT ADH ALA-127.
RX MEDLINE=95179179; PubMed=7874174;
RA Pollak M.R., Brown E.M., Estep H.L., McLaine P.N., Kifor O., Park J.,
  Hebert S.C., Seidman C.E., Seidman J.G.;
RT "Autosomal dominant hypercalcaemia caused by a Ca(2+)-sensing receptor
  gene mutation.";
RL Nat. Genet. 8:303-307(1994).
[8]
RP VARIANTS FHH MET-62; CYS-66; MET-138; GLU-143 AND GLN-227.
RX MEDLINE=95243222; PubMed=7726161;
RA Chou Y.-H.W., Pollak M.R., Brandi M.L., Toss G., Arnqvist H.,
  Atkinson A.B., Papapoulos S.E., Marx S., Brown E.M., Seidman J.G.,
  Seidman C.E.;
RT "Mutations in the human Ca(2+)-sensing-receptor gene that cause
  familial hypocalciuric hypercalcaemia";
RL Am. J. Hum. Genet. 56:1075-1079(1995).
[9]
RP SEQUENCE OF 1-61 FROM N.A., AND VARIANT FHH ALA-39.
RX MEDLINE=95403641; PubMed=7673400;
RA Aida K., Koishi S., Inoue M., Nakazato M., Tawata M., Onaya T.;
RT "Familial hypocalciuric hypercalcaemia associated with mutation in the
  human Ca(2+)-sensing receptor gene.";
RL J. Clin. Endocrinol. Metab. 80:2594-2598(1995).
[10]
RP VARIANTS NSHPT LEU-227 AND TYR-598.
RX MEDLINE=96292293; PubMed=8675635;
RA Pearce S.H.S., Trump D., Wooding C., Besser G.M., Chew S.L.,
  Grant D.B., Heath D.A., Hughes I.A., Paterson C.R., Whyte M.P.,
  Thakker R.V.;
RT "Calcium-sensing receptor mutations in familial benign hypercalcaemia
  and neonatal hyperparathyroidism.";
RL J. Clin. Invest. 96:2683-2692(1995).
[11]
RP VARIANTS ADHP THR-116; HIS-681 AND SER-806, AND VARIANT SER-851.
RX MEDLINE=96311554; PubMed=8733126;
RA Baron J., Winer K.K., Yanovski J.A., Cunningham A.W., Laue L.,
  Zimmerman D., Cutler G.B. Jr.;
RT "Mutations in the Ca(2+)-sensing receptor gene cause autosomal
  dominant and sporadic hypoparathyroidism.";
RL Hum. Mol. Genet. 5:601-606(1996).
[12]
RP VARIANT FHH ARG-174.
RX MEDLINE=97442275; PubMed=9298824;
RA Ward B.K., Stuckey B.G.A., Gutteridge D.H., Laing N.G., Pullan P.T.,
  Ratajczak T.;
RT "A novel mutation (L174R) in the Ca2+-sensing receptor gene
  associated with familial hypocalciuric hypercalcaemia.";
RL Hum. Mutat. 10:233-235(1997).
```

[13]  
 RN VARIANT FHH GLU-557.  
 RP MEDLINE=21603857; PubMed=11762699;  
 RA Nakayama T., Minato M., Nakagawa M., Soma M., Tobe H., Aoi N.,  
 RA Kosuge K., Sato M., Ozawa Y., Kamatsuse K., Kokubun S.;  
 RT "A novel mutation in Ca<sup>2+</sup>-sensing receptor gene in familial  
 RT hypocalcemic hypercalcemia.";  
 RL Endocrine 15:277-282(2001).  
 CC -!- FUNCTION: SENSE CHANGES IN THE EXTRACELLULAR CONCENTRATION OF  
 CC CALCIUM IONS. THE ACTIVITY OF THIS RECEPTOR IS MEDIATED BY A  
 CC G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-CALCIUM SECOND  
 CC MESSENGER SYSTEM.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
 CC produced by alternative splicing.  
 CC -!- TISSUE SPECIFICITY: FOUND IN KIDNEY, BUT NOT IN BRAIN, LUNG,  
 CC LIVER, HEART, SKELETAL MUSCLE, OR PLACENTA.  
 CC -!- DISEASE: DEFECTS IN CASR ARE A CAUSE OF FAMILIAL HYPOCALCAEMIC  
 CC HYPERCALCAEMIA (FHH) AND NEONATAL SEVERE HYPERPARATHYROIDISM  
 CC (NSHPT). TWO INHERITED CONDITIONS CHARACTERIZED BY ALTERED CALCIUM  
 CC HOMEOSTASIS. THE MUTATIONS REDUCE THE ACTIVITY OF THE RECEPTOR.  
 CC FHH AFFECTED INDIVIDUALS EXHIBIT MILD OR MODEST HYPERCALCAEMIA,  
 CC RELATIVE HYPOCALCAEMIA, AND INAPPROPRIATELY NORMAL PTH LEVELS. IN  
 CC CONTRAST, NSHPT IS A RARE AUTOSOMAL RECESSIVE LIFE-THREATENING  
 CC DISORDER CHARACTERIZED BY VERY HIGH SERUM CALCIUM CONCENTRATIONS,  
 CC SKELETAL DEMINERALIZATION, AND PARATHYROID HYPERPLASIA. IN SOME  
 CC INSTANCES NSHPT HAS BEEN DEMONSTRATED TO BE THE HOMOZYGOUS FORM OF  
 CC FHH.  
 CC -!- DISEASE: DEFECTS IN CASR ARE ALSO THE CAUSE OF AUTOSOMAL DOMINANT  
 CC HYPOCALCAEMIA (ADH) IN WHICH THE RECEPTOR IS ACTIVATED AT SUBNORMAL  
 CC CA(2+) LEVELS.  
 CC -!- DISEASE: DEFECTS IN CASR ARE ALSO THE CAUSE OF AUTOSOMAL DOMINANT  
 CC HYPOPARATHYROIDISM (ADHP). ADHP IS CHARACTERIZED BY HYPOCALCAEMIA  
 CC AND HYPERPHOSPHATEMIA DUE TO INADEQUATE SECRETION OF PARATHYROID  
 CC HORMONE. SYMPTOMS ARE SEIZURES, TETANY AND CRAMPS.  
 CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X81086; CAA56990.1; -  
 CC EMBL; U20759; AAA86503.1; -  
 CC EMBL; U20760; AAA86504.1; -  
 CC EMBL; D50855; BAA09453.1; -  
 CC EMBL; S83176; AAB46873.1; -  
 CC EMBL; S79217; AAB35262.2; -  
 CC EMBL; S68032; AAB29413.2; ALT\_SEQ.  
 CC EMBL; S68033; AAB29414.1; -  
 CC EMBL; S68036; AAB29415.1; -  
 CC EMBL; S81755; AAD14370.1; -  
 CC Genew; HGNC:1514; CASR.  
 CC MIM; 601199; -  
 CC MIM; 145980; -  
 CC MIM; 239200; -  
 CC MIM; 601198; -  
 CC InterPro; IPR001828; ANF\_receptor.  
 CC InterPro; IPR000337; GPCR\_Mgr.  
 CC Pfam; PF00003; 7tm\_3; 1.  
 CC Pfam; PF01094; ANF\_receptor; 1.  
 CC PRINTS; PR00248; GPCRMRG.  
 CC PROSITE; PS00979; G-PROTEIN\_RECEP\_F3\_1; 1.  
 CC PROSITE; PS00980; G-PROTEIN\_RECEP\_F3\_2; 1.  
 CC PROSITE; PS00981; G-PROTEIN\_RECEP\_F3\_3; 1.  
 CC PROSITE; PS00259; G-PROTEIN\_RECEP\_F3\_4; 1.  
 CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;  
 KW Disease mutation; Alternative splicing; Polymorphism.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1078  
 FT EXTRACELLULAR CALCIUM-SENSING RECEPTOR.

FT DOMAIN 20 612 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 613 635 I (POTENTIAL).  
 FT DOMAIN 636 649 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 650 670 II (POTENTIAL).  
 FT DOMAIN 671 681 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 682 700 III (POTENTIAL).  
 FT DOMAIN 701 724 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 725 745 IV (POTENTIAL).  
 FT DOMAIN 746 769 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 770 792 V (POTENTIAL).  
 FT DOMAIN 793 805 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 806 828 VI (POTENTIAL).  
 FT DOMAIN 829 836 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 837 862 VII (POTENTIAL).  
 FT DOMAIN 863 1078 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 130 130 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 287 287 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 386 386 N-LINKED (GLCNAC. .) (POTENTIAL).  
 Query Match 4.0%; Score 127; DB 1; Length 1078;  
 Best Local Similarity 22.5%; Pred. No. 0.072;  
 Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;  
 QY 311 OKVIDIDQSSDQWFSSTHKC---HLNNE-CMPKIGLGVLGAYECICKAG---FVHPGV 363  
 DB 475 EQVTFDEC---GDLVGNYSIINWHLSPEDGSIIVFEVGY---YNYAKKGERLFINEEK 527  
 QY 364 LPVNNFRRRGPDQH-----ISGSTRDVSPD-----AYVCIPCREG-----CPFCAD 404  
 DB 528 ILWSGFSREVPNSCRDCLAGTRGILIEGPTCCFCEVCEPDGEVSDETDASACKNCPD 587  
 QY 405 D-----SPCEVQEDKYLR-----LAIISQGLCMLDFVSMVLVYVHFRKAKSIRASG 451  
 DB 588 DFWSNENHTSCIAKEIEFLSWTEPFGLIALTFAVLGIFLTAFVLGVFKFRNTPVVKATN 647  
 QY 452 LILLETILFGLLLVPVILYFEPSTPRCLLWRLIGFATVYGVTVTLKLRVVKVFL 511  
 DB 648 RELSVLLLFSLCCSSSLFFTFGEQDWTQRURQAFGISFVLCISCLVTKNRVLLVFE 707  
 QY 512 SRTAQRIPTYMTGGRYMRMLAV-----ILLVVFVFLIGWTSV-CQNLKQISLI----- 559  
 DB 708 AKIPTSFHRKMWGLNLQFLFLCTFMQIVICVILWYATPPSSYRNQOELEDEIITTCHE 767  
 QY 560 GQGTKSDHLIFNMCLI 575  
 DB 768 GSLMALGFLIGYTCLL 783  
 RESULT 6  
 CASR\_BOVIN STANDARD; PRT; 1085 AA.  
 ID CASR\_BOVIN  
 AC P35384;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Extracellular calcium-sensing receptor precursor (CaSR) (Parathyroid  
 DE Cell calcium-sensing receptor).  
 GN CASR OR GPRC2A OR PCAR1.  
 OS Bos taurus (Bovine).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 CC Bovidae; Bovinae; Bos.  
 CC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Parathyroid;  
 RX MEDLINE=94077182; PubMed=8255296;  
 RA Brown E.M., Gamba G., Riccardi D., Lombardi M., Butters R., Kifor O.,  
 RA Sun A., Hediger M.A., Lyttton J., Hebert S.C.;  
 RT "Cloning and characterization of an extracellular Ca(2+)-sensing  
 RT receptor from bovine parathyroid.";





```
KW G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family.
FT SIGNAL 1 18
FT CHAIN 19 872
FT DOMAIN 19 567
FT TRANSMEM 568 590
FT DOMAIN 591 604
FT TRANSMEM 605 625
FT DOMAIN 626 636
FT TRANSMEM 637 655
FT DOMAIN 656 679
FT TRANSMEM 680 700
FT DOMAIN 701 725
FT TRANSMEM 726 747
FT DOMAIN 748 760
FT TRANSMEM 761 783
FT DOMAIN 784 793
FT TRANSMEM 794 819
FT DOMAIN 820 872
FT CARBOHYD 203 203
FT CARBOHYD 286 286
FT CARBOHYD 338 338
FT CARBOHYD 402 402
FT CARBOHYD 547 547
FT CONFLICT 12 12
FT CONFLICT 210 210
FT CONFLICT 496 496
FT CONFLICT 748 748
FT CONFLICT 776 776
FT SEQUENCE 872 AA; 95507 MW; 058608C35C701E9D CRC64;

Query Match
Best Local Similarity 3.8%; Score 120; DB 1; Length 872;
Matches 49; Conservative 28; Mismatches 70; Indels 40; Gaps 9;

QY 390 YVCLPCREGCPF-----CADD-----SPCFVQEDKYLR-----LAITS 422
Db 520 WLICPCQ---PYEYRLDEFTCADGLGWPNASLTGCFELPQEIYRWDAWAGPVTIAC 576
QY 423 FQGLCLMLDFVSMVYVYHFKAKSIRASGLILLETILFGLSLLLYFPVILYEPSTFCI 482
Db 577 LGALATL--FVLGVFRH--NATPVVKASGRELYLLGGVFLCYMTFFIAKPSAVCT 633
QY 483 LIRWARLLGFATVIGVTVLKLRHLKVF--LSRTAQRIPYMT--GGVRMRLAVI-----LL 535
Db 634 LRRLLGLTAFSCYCSALLTKTNRIARIFGAREGAORPRFISPASQVATCLALISQQLI 693
QY 536 VVFWFLI 542
Db 694 VVAVLVV 700

RESULT 8
MGR2_RAT STANDARD; PRT; 872 AA.
AC P31421;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metabotropic glutamate receptor 2 precursor (mGluR2).
GN GRM2 OR GPRC1B OR MGLUR2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92110002; PubMed=1309649;
RA Tanabe Y., Masu M., Ishii T., Shigemoto R., Nakanishi S.;
RT "A family of metabotropic glutamate receptors.";
RL Neuron 8:169-179(1992).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS
```

```
CC CC MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE ACTIVITY.
CC CC MAY MEDIANE SUPPRESSION OF NEUROTRANSMISSION OR MAY BE INVOLVED IN
CC CC SYNAPTOGENESIS OR SYNAPTIC STABILIZATION.
CC CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC CC -!- TISSUE SPECIFICITY: IS WIDELY DISTRIBUTED IN THE CNS AND PROMINENT
CC CC EXPRESSION IS SEEN IN GOLGI CELLS OF THE CEREBELLUM AND SOME
CC CC PARTICULAR NEURONAL CELLS IN OTHER BRAIN REGIONS.
CC CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC CC STRONGEST, TO MGLUR3.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; M92075; -; NOT_ANNOTATED_CDS.
CC CC PIR; JH0561; JH0561-
CC CC InterPro; IPR001828; ANF_receptor.
CC CC InterPro; IPR000337; GPCR_Mgr.
CC CC Pfam; PF00003; 7tm_3; 1.
CC CC Pfam; PF01094; ANF_receptor; 1.
CC CC PRINTS; PR00248; GPCRMR.
CC CC PROSITE; PS00979; G_PROTEIN_RECP_F3_1; 1.
CC CC PROSITE; PS00980; G_PROTEIN_RECP_F3_2; 1.
CC CC PROSITE; PS00981; G_PROTEIN_RECP_F3_3; 1.
CC CC PROSITE; PS00259; G_PROTEIN_RECP_F3_4; 1.
CC CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW Multigene family.
FT SIGNAL 1 18
FT CHAIN 19 872
FT DOMAIN 19 567
FT TRANSMEM 568 590
FT DOMAIN 591 604
FT TRANSMEM 605 625
FT DOMAIN 626 636
FT TRANSMEM 637 655
FT DOMAIN 656 679
FT TRANSMEM 680 700
FT DOMAIN 701 725
FT TRANSMEM 726 747
FT DOMAIN 748 760
FT TRANSMEM 761 783
FT DOMAIN 784 793
FT TRANSMEM 794 819
FT DOMAIN 820 872
FT CARBOHYD 203 203
FT CARBOHYD 286 286
FT CARBOHYD 338 338
FT CARBOHYD 402 402
FT CARBOHYD 547 547
FT SEQUENCE 872 AA; 95773 MW; 1E74CABD6AD4BED9 CRC64;

Query Match
Best Local Similarity 3.6%; Score 116; DB 1; Length 872;
Matches 93; Conservative 63; Mismatches 184; Indels 152; Gaps 22;

QY 141 QSNKSGRE-----QNLQDDLDWYQALVWSLLE----GEPSISRAAITFSDLSAPAPQVF 191
Db 271 RSEDARELLAATQRLNASFTWVSDGWGALESVWAGSERAEGAITIELASYPISDFASY 330
QY 192 LOATREESRILLQDSSSAPHLANATLETWFPHGLRRKWRPHLRRGNQGPGRGLGHSWR 251
Db 331 FQS-----LDPWNNS-----RNPWF---REFWEERFH-----CSFR 358
QY 252 RKDGLGGDKSHFKWSPPYLECEGNGSKYKPGWLVTLSAIVGLQNLVLP----- 298
Db 359 QRD-----CAHSLRAVPFQESKIMFVNAVYAMAHALHNHRLCPNTHLCDAMRPVN 414
QY 299 -----EFRGNKVDINLQKVDI-DQCSDDGWFSGTHKCHLNNSCMPKIGLGFVLGAYE 351
```



Db 415 GRLLKDFVNLVKEDAPRRPADTDEVRDFRGGD-----IGRYN 454

Qy 352 CIC-----KAGFYHPGVLPVNNF-----RRRGPDQHSSTKDVSE 387

Db 455 IFTYLRAGSGRYRYKGVYWAEGTLDFSPWASPSAGPLPASCSEPCQLONEVKSQVP 514

Qy 388 E---AYVCLPREGCPF-----CADD-----SPCFVQEDKYL- 417

Db 515 GEVCCWLCIPQ---PYEYRLDEFTCADGGLGYWPNASLTGCFELPQBYIRWGDAWAGVP 571

Qy 418 LAIISFQGLCMLDFVSMVYVYHFRKAKSIRASGLILLETILFGLSLLLYFPVILYEPSP 477

Db 572 VTIACLGALATL--FVLGVFVRH--NATPVVKASGRELCYILLGGVFLCYCMTFVFIKAPS 628

Qy 478 TFRGILLRWARLLGFATVIGTVTLKLRVLKVF--LSRTAQIRPYMT--GGRVWRMLAVI- 533

Db 629 TAVCTLRRLGLTAFSCVCSALLTKTNRIARIFGARGAORPREISPASQVAICLALIS 688

Qy 534 ---LLVWFVFLI 542

Db 689 GQLLIVRAWLVV 700

RESULT 9

MGRI\_HUMAN

ID MGRI\_HUMAN STANDARD; PRT; 1194 AA.

AC Q13255; Q14757; Q14758; Q9UGT0; Q9UGS9; Q9NU10;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Metabotropic glutamate receptor 1 precursor (mglur1).

GN GRI1 OR GPRC1A OR MGLUR1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

OX NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.

RP MEDLINE=9602774; PubMed=7476890;

RA Desai M.A., Burnett J.P., Mayne N.G., Schoepp D.D.;

RT "Cloning and expression of a human metabotropic glutamate receptor 1

RT alpha: enhanced coupling on co-transfection with a glutamate

RT transporter.";

RL Mol. Pharmacol. 48:648-657(1995).

[2]

RP SEQUENCE FROM N.A.

RP MEDLINE=97231349; PubMed=9076744;

RA Stephan D., Bon C., Holzwarth J.A., Galvan M., Pruss R.M.;

RT "Human metabotropic glutamate receptor 1: mRNA distribution,

RT chromosome localization and functional expression of two splice

RT variants.";

RL Neuropharmacology 35:1649-1660(1996).

[3]

RP SEQUENCE FROM N.A.

RA Bates K.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR IS

CC MEDIATED BY A G-PROTEIN THAT ACTIVATES A PHOSPHATIDYLINOSITOL-

CC CALCIUM SECOND MESSENGER SYSTEM. MAY PARTICIPATE IN THE CENTRAL

CC ACTION OF GLUTAMATE IN THE CNS, SUCH AS LONG-TERM POTENTIATION IN

CC THE HIPPOCAMPUS AND LONG-TERM DEPRESSION IN THE CEREBELLUM.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA;

CC ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.

CC STRONGEST, TO MGLUR5.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL; U31215; AAA87843.1; -

CC EMBL; U31216; AAA87844.1; -

DR EMBL; L76627; AAB05337.1; -

DR EMBL; L76631; AAB05338.1; -

DR EMBL; AL096867; CAB75694.1; -

DR EMBL; AL035698; CAB65991.1; -

DR EMBL; AL035698; CAB65992.1; -

DR Genew; HGNC:4593; GRM1.

DR MIM; 604473; -

DR InterPro; IPR001828; ANF\_receptor.

DR InterPro; IPR000337; GPCR\_Mgr.

DR Pfam; PF00003; 7tm\_3; 1.

DR Pfam; PF01094; ANF\_receptor; 1.

DR PRINTS; PR00248; GPCRMR.

DR PROSITE; PS00979; G\_PROTEIN\_RECEP\_F3\_1; 1.

DR PROSITE; PS00980; G\_PROTEIN\_RECEP\_F3\_2; 1.

DR PROSITE; PS00981; G\_PROTEIN\_RECEP\_F3\_3; 1.

DR PROSITE; PS0259; G\_PROTEIN\_RECEP\_F3\_4; 1.

DR G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;

KW Multigene family; Alternative splicing.

KW SIGNAL 1..18

FT CHAIN 19 1194 METABOTROPIC GLUTAMATE RECEPTOR 1.

FT DOMAIN 19 592 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 593 615 I (POTENTIAL).

FT DOMAIN 616 629 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 630 650 II (POTENTIAL).

FT DOMAIN 651 661 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 662 680 III (POTENTIAL).

FT DOMAIN 681 706 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 707 727 IV (POTENTIAL).

FT DOMAIN 728 750 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 751 772 V (POTENTIAL).

FT TRANSMEM 773 785 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 786 808 VI (POTENTIAL).

FT TRANSMEM 809 814 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 815 840 VII (POTENTIAL).

FT DOMAIN 841 1194 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 1014 1035 GLN/PRO-RICH.

FT DOMAIN 1067 1081 GLN/PRO-RICH.

FT TRANSMEM 1095 1130 ASP/GLU-RICH (ACIDIC).

FT DOMAIN 1142 1194 SER-RICH.

FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 515 515 N-LINKED (GLCNAC. .) (POTENTIAL).

FT VARSPLIC 887 906 NSNGKSVSWSEPGGQVPGK -> KKQPEFSPSQCPSAH

FT VARSPLIC 907 1194 VQL (IN ISOFORM BETA).

FT CONFLICT 593 593 MISSING (IN ISOFORM BETA).

FT CONFLICT 993 993 P -> S (IN REF. 2 AND 3).

FT CONFLICT 993 993 P -> S (IN REF. 3).

SQ SEQUENCE 1194 AA; 132376 MW; 970E5IAF40584F40 CRC64;

Query Match 3.6%; Score 116; DB 1; Length 1194;

Best Local Similarity 22.1%; Pred. No. 0.59;

Matches 49; Conservative 34; Mismatches 95; Indels 44; Gaps 6;

Qy 357 GFYPHGVLPVNNFRRRDPDQHSSTKDVSE-----AVVCLPREG- 398

Db 498 GTWHEGVNLIDDDYIQ---MNKSGVRSVCEPCLKGQIKVIRKGEVSCCWTCTACKENE 554

Qy 399 -----CPFC-----ADDSFCVQEDKYL------LAIISFQGLCMLDFVSMV 438

Db 555 YVQDEFTCKACDLGWNPNADTGCETPIPVRYLEWSNTEPIIAIFASCGILVTLFVTLIF 614

Qy 439 YHFRKAKSIRASGLILLETILFGLSLLLYFPVILYFPPFCRILLRWARLLGFATVYGT 498

Db 615 VLYRDTVPVKSRRSRELCYIILAGILGVCPFTLIARPTTSCYLORLLVGLSSAMCYSA 674

Qy 499 VTLKLRVLYKVF---SRTAQIRPYMTGGRVWRMLAVILV 537

Db 675 LVTKTNRIARILAGSKKIKCTRKPRFNSAWAQVITIASILISV 716

```
RESULT 10
MGR3_HUMAN
ID MGR3_HUMAN STANDARD; PRT; 877 AA.
AC Q14832;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Metabotropic glutamate receptor 3 precursor (mGluR3).
GN GRM3 OR GPRC1A OR MGLUR3.
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
PC TISSUE=Brain;
MEDLINE=96437205; PubMed=8840013;
MAKOFF A., VOLPE F., LEICHUK R., HARRINGTON K., EMSON P.;
"molecular characterization and localization of human metabotropic
glutamate receptor type 3.";
RL Brain Res. Mol. Brain Res. 40:55-63(1996).
CC -!- FUNCTION: RECEPTOR FOR GLUTAMATE. THE ACTIVITY OF THIS RECEPTOR
CC IS MEDIATED BY A G-PROTEIN THAT INHIBITS ADENYLATE CYCLASE
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO FAMILY 3 OF G-PROTEIN COUPLED RECEPTORS.
CC STRONGEST, TO MGLUR2.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X77748; CAA54796.1; -.
CC Genew, HGNC:4595; GRM3.
CC MIM: 601115; -.
CC InterPro: IPR001828; ANF_receptor.
CC InterPro: IPR000337; GPCR_Mgr.
CC Pfam: PF00003; 7tm.3; 1.
CC Pfam: PF01094; ANF_receptor; 1.
CC PRINTS: PR00248; GPCR_MGR.
CC PROSITE: PS00979; G_PROTEIN_REC_F3_1; 1.
CC PROSITE: PS00980; G_PROTEIN_REC_F3_2; 1.
CC PROSITE: PS00981; G_PROTEIN_REC_F3_3; 1.
CC PROSITE: PS0259; G_PROTEIN_REC_F3_4; 1.
CC G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
CC Multigene family.
CC -----
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 877 METABOTROPIC GLUTAMATE RECEPTOR 3.
FT DOMAIN 21 574 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 575 597 I (POTENTIAL).
FT DOMAIN 598 611 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 612 632 II (POTENTIAL).
FT DOMAIN 633 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 662 III (POTENTIAL).
FT DOMAIN 663 686 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 687 707 IV (POTENTIAL).
FT DOMAIN 708 732 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 733 754 V (POTENTIAL).
FT DOMAIN 755 767 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 768 790 VI (POTENTIAL).
FT DOMAIN 791 800 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 801 826 VII (POTENTIAL).
FT DOMAIN 827 877 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 207 207 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 412 412 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC... ) (POTENTIAL).
```

```
SQ SEQUENCE 877 AA; 98619 MW; 66F28663CE35F740 CRC64;
Query Match 3.6%; Score 115.5; DB 1; Length 877;
Best Local Similarity 18.6%; Pred. NO. 0.43;
Matches 90; Conservative 72; Mismatches 196; Indels 125; Gaps 23;
QY 138 VMLQSNKRSRE-----QNLQDDLDWYQALVW----SLLEGEPSISRAAITFTDLSAPAP 188
D 138 VMLQSNKRSRE-----QNLQDDLDWYQALVW----SLLEGEPSISRAAITFTDLSAPAP 188
DB 272 LFMRSDDSRILAAASRANASFTWVSDGWAQAEIITKSGSEHVAYGAILTELAS----- 325
QY 189 QVFLQATREESRIILLQDLSSAPHLANATLETWFHGLRKRWRPHLHRRGNQGRGLGH 248
D 189 QVFLQATREESRIILLQDLSSAPHLANATLETWFHGLRKRWRPHLHRRGNQGRGLGH 248
DB 326 ---QPVROQFDR-VFQSLNPYNH-----RNPWRDF----- 352
QY 249 SWRRK---DGLGGDKSHFKWSPPYLECENGSKYK-----GWLVTLSAIYGLQPNLVP 298
D 249 SWRRK---DGLGGDKSHFKWSPPYLECENGSKYK-----GWLVTLSAIYGLQPNLVP 298
DB 353 -WEQFOCSLQNKRNRRVCDKHLAIDSSNYEQESKIMFVVVNAVYAMAHALHKWRTLCP 411
QY 299 ---EFRGMVKVDINLQKVDIDOCSSDGFSGTHKCHLNNSCEMPIKGLGFLVGLYAYECI-- 353
D 299 ---EFRGMVKVDINLQKVDIDOCSSDGFSGTHKCHLNNSCEMPIKGLGFLVGLYAYECI-- 353
DB 412 NTKLCDAMKI-LDGKKLYDKYLLKIN-FTAPFPNPKDADSIVKFDTFDGMGRYNNVFN 469
QY 354 -----CKAGFYHFGV-LPVNMF---RRRGPDQHIS-----GSTKDVSEE---AYVC 392
D 354 -----CKAGFYHFGV-LPVNMF---RRRGPDQHIS-----GSTKDVSEE---AYVC 392
DB 470 QNVGGKYSYLKVGHWAEITSLDVNSIHWSRNSVPTSOCDPCAPNEMKMQPDVCCWIC 529
QY 393 LPCR-----CCPFC-----ADDSPCFVOEDKYLR-----LAISFQGL 426
D 393 LPCR-----CCPFC-----ADDSPCFVOEDKYLR-----LAISFQGL 426
DB 530 IPCEPYEYLADFTCMDCGSGQWPTADLTGCTDLPEDYIRWEDAWAIGPVTIACLGFMCT 589
QY 427 CMLLDFVSLVYVYHFRKAKSTRASGLILLETILFGLSLLYFPVVLVYFEPSTFCILLRW 486
D 427 CMLLDFVSLVYVYHFRKAKSTRASGLILLETILFGLSLLYFPVVLVYFEPSTFCILLRW 486
DB 590 CMV-----VTVFKHNTPLVKASGRELCYLLFGVGLSYCMCTFFFTAKSPVICALRRL 644
QY 487 ARLLGFATVGTGVTFLKLRHLKVF--LSRTAQRIPYMT-GGRVMRLAVIL----LVVFW 539
D 487 ARLLGFATVGTGVTFLKLRHLKVF--LSRTAQRIPYMT-GGRVMRLAVIL----LVVFW 539
DB 645 GLGSSFAICYSALLTKTNCIARIIDFGVKNGAQRKPFISPSQVFCILGLLILVQIVMYSVW 704
QY 540 FLI 542
D 705 LIL 707
RESULT 11
MGR1_RAT
ID MGR1_RAT STANDARD; PRT; 1199 AA.
AC P23385;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metabotropic glutamate receptor 1 precursor (mGluR1).
GN GRM1 OR GPRC1A OR MGLUR1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=91156047; PubMed=1847995;
RA Masu M., Tanabe Y., Tsuhida K., Shigemoto R., Nakanishi S.;
RT "Sequence and expression of a metabotropic glutamate receptor.";
RL Nature 349:760-765(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92022526; PubMed=1656524;
RA Houamed K.M., Kuiper J.L., Gilbert T.L., Haldeman B.A., O'Hara P.J.,
RA Mulvihill E.R., Almers W., Hagen F.S.;
RT "Cloning, expression, and gene structure of a G protein-coupled
RT glutamate receptor from rat brain.";
RL Science 252:1318-1321(1991).
RN [3]
```



DR	PIR; JH0562; JH0562.	
DR	InterPro: IPR001828; ANF_receptor.	
DR	InterPro: IPR000337; GPCR_Mgr.	
DR	pfam: PF00003; 7tm_3; 1.	
DR	Pfam: PF01094; ANF_receptor; 1.	
DR	PRINTS: PR00248; GPCRMR.	
DR	PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.	
DR	PROSITE; PS00980; G_PROTEIN_RECEP_F3_2; 1.	
DR	PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.	
DR	PROSITE; PS00259; G_PROTEIN_RECEP_F3_4; 1.	
KW	G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;	
FT	Multigene family.	
FT	1	22
FT	SIGNAL	POTENTIAL.



Db 382 SSLLTRHHVQDSKMGFVINALIYSMAYGLHNMOMSLCPGYAGLCDAMKPIDGRKLLDSLM 441  
 QY 312 KVIDOCSSDGFSGTHKCHLNNSCMPYKGLGFLVGLGAYECICKAGFYHPGLPV----- 366  
 Db 442 KTNFTGVSGMILFDGSDGPRGYELNFKEMG---KDFDYNVGSWDNGELKMDDEV 498  
 QY 367 -----NNFRRRGPDQHT-SGSTKDVSSE---AYVCLPCRE-----CPFC-----A 403  
 Db 499 WSKKNLIIRSVCEPCEKGOIKVIRKEVSCCWCTCPCKENEVEFDEYCKACQLGSWPT 558  
 QY 404 DD-SPCFVQEDKYLR-----LAISFGQLCMLLDFVSMVYVHFRKAKSIRASGLILLE 456  
 Db 559 DDLTGCDLIPVQYLRWGDPEPIAAVFAACGLIATLFTVFIITYRDTVPWAKSSSRELCY 618  
 QY 457 TILFGSLLYFPVILYFEPSTFRCLLRWRLGLFATVGVTVTLKHLRVLKVFLSRTAQ 516  
 Db 619 IILAGICLGLYCTFLCIAPKQIYCVLQIRIGIGLSPAMYSALVTKNTNRIARI-LAGSKK 677  
 Db 517 RI-----PYMTGGRVMRLAVILLV 537  
 Db 678 KICTKKPRFMSACAQLVIAFILICI 702  
 NCBI\_TaxID=10116;  
 RESULT 15  
 NR2\_RAT  
 ID NR2\_RAT STANDARD; PRT; 868 AA.  
 AC O35569; O35570; O35571; O35572; O35073;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pro-neuregulin-2 precursor (Pro-NRG2) [Contains: Neuregulin-2 (NRG-2)  
 DE (Neural- and thymus-derived activator for ERBB kinases) (NTAK)].  
 GN NR2 OR NTAK.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RX MEDLINE-98006324; PubMed-9348101;  
 RA Higashiyama S., Horikawa M., Yamada K., Ichino N., Nakano N.,  
 RA Nakagawa T., Miyagawa J., Matsushita N., Nagatsu T., Taniguchi N.,  
 RA Ishiguro H.;  
 RT "A novel brain-derived member of the epidermal growth factor family  
 RT that interacts with ErbB3 and ErbB4.";  
 RL J. Biochem. 122:675-680(1997).  
 [2]  
 SEQUENCE OF 109-868 FROM N.A. (ISOFORMS NRG2-ALPHA AND NRG2-BETA).  
 TISSUE=Cerebellum;  
 RX MEDLINE-97311397; PubMed-9168114;  
 RA Chang H., Riese D.J. II, Gilbert W., Stern D.F., McMahon U.J.;  
 RT "Ligands for ErbB-family receptors encoded by a neuregulin-like  
 RT gene.";  
 RL Nature 387:509-512(1997).  
 CC -!- FUNCTION: DIRECT LIGAND FOR ERBB3 AND ERBB4 TYROSINE KINASE  
 CC RECEPTORS. CONCOMITANTLY RECRUITS ERBB1 AND ERBB2 CORECEPTORS,  
 CC RESULTING IN LIGAND-STIMULATED TYROSINE PHOSPHORYLATION AND  
 CC ACTIVATION OF THE ERBB RECEPTORS. MAY ALSO PROMOTE THE  
 CC HETERODIMERIZATION WITH THE EGF RECEPTOR.  
 CC -!- SUBCELLULAR LOCATION: EXISTS AS AN TYPE I MEMBRANE PROTEIN AND AS  
 CC A PROTEOLYTICALLY RELEASED SOLUBLE GROWTH FACTOR FORM. THE  
 CC MEMBRANE-BOUND FORM DOES NOT SEEM TO BE ACTIVE (BY SIMILARITY).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS: NTAK-ALPHA1 (SHOWN  
 CC HERE), NTAK-ALPHA2A, NTAK-ALPHA2B/NTAK-ALPHA2-1P, NTAK-BETA, NTAK-  
 CC GAMMA, NRG2-ALPHA AND NRG2-BETA; ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING. THE ALPHA-TYPE AND BETA-TYPE DIFFER IN THE EGF-LIKE  
 CC DOMAIN.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST PARTS OF THE BRAIN,  
 CC ESPECIALLY THE OLFACTORY BULB AND CEREBELLUM WHERE IT LOCALIZES IN  
 CC GRANULE AND PURKINJE CELLS. IN THE HIPPOCAMPUS, FOUND IN THE  
 CC GRANULE CELLS OF THE DENTATE GYRUS. IN THE BASAL FOREBRAIN, FOUND  
 CC IN THE CHOLINERGIC CELLS. IN THE HINDBRAIN, WEAKLY DETECTABLE IN  
 CC THE MOTOR TRIGEMINAL NUCLEUS. NOT DETECTED IN THE HYPOTHALAMUS.

CC ALSO FOUND IN THE LIVER AND IN THE THYMUS. NOT DETECTED IN HEART,  
 CC ADRENAL GLAND, OR TESTIS.  
 CC -!- DEVELOPMENTAL STAGE: IN THE EMBRYO, EXPRESSED IN THE BRAIN OF  
 CC E11.5 EMBRYOS WHERE IT IS FOUND IN THE TELENCEPHALON, BUT NOT IN  
 CC THE HINDBRAIN. NOT FOUND IN THE HEART. IN THE ADULT, FOUND IN  
 CC BRAIN AND THYMUS.  
 CC -!- DOMAIN: THE CYTOPLASMIC DOMAIN MAY BE INVOLVED IN THE REGULATION  
 CC OF TRAFFICKING AND PROTEOLYTIC PROCESSING. REGULATION OF THE  
 CC PROTEOLYTIC PROCESSING INVOLVES INITIAL INTRACELLULAR DOMAIN  
 CC DIMERIZATION (BY SIMILARITY).  
 CC -!- DOMAIN: ERBB RECEPTOR BINDING IS ELICITED ENTIRELY BY THE EGF-LIKE  
 CC DOMAIN (BY SIMILARITY).  
 CC -!- PTM: PROTEOLYTIC CLEAVAGE CLOSE TO THE PLASMA MEMBRANE ON THE  
 CC EXTERNAL FACE LEADS TO THE RELEASE OF THE SOLUBLE GROWTH FACTOR  
 CC FORM (BY SIMILARITY).  
 CC -!- PTM: EXTENSIVE GLYCOSYLATION PRECEDES THE PROTEOLYTIC CLEAVAGE (BY  
 CC SIMILARITY).  
 CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
 CC -!- SIMILARITY: BELONGS TO THE NEUREGULIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D89995; BAA23344.1; -;  
 CC EMBL; D89996; BAA23345.1; -;  
 CC EMBL; D89997; BAA23346.1; -;  
 CC EMBL; D89998; BAA23347.1; -;  
 CC EMBL; AB001576; BAA23348.1; -;  
 CC HSSP; Q12784; IHRE.  
 CC InterPro; IPR000561; EGF-like.  
 CC InterPro; IPR003006; Ig\_MHC.  
 CC InterPro; IPR003598; Ig\_C2.  
 CC InterPro; IPR002154; Neuregulin.  
 CC Pfam; PF00008; EGF; 1.  
 CC Pfam; PF00047; Ig; 1.  
 CC Pfam; PF02158; Neuregulin; 1.  
 CC SMART; SM00181; EGF; 1.  
 CC SMART; SM00408; IGG2; 1.  
 CC PROSITE; PS00022; EGF\_1; 1.  
 CC PROSITE; PS01186; EGF\_2; 1.  
 KW Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein;  
 KW Transmembrane; Multigene family; Alternative splicing.  
 FT PROPEP 1 127  
 FT CHAIN 128 868 PRO-NEUREGULIN-2, MEMBRANE-BOUND FORM.  
 FT CHAIN 128 428 NEUREGULIN-2.  
 FT DOMAIN 128 429 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 430 450 INTERNAL SIGNAL SEQUENCE (POTENTIAL).  
 FT DOMAIN 451 868 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 266 334 IG-LIKE C2-TYPE DOMAIN.  
 FT DOMAIN 346 356 SER/THR-RICH.  
 FT DOMAIN 357 398 EGF-LIKE.  
 FT DOMAIN 22 32 POLY-SER.  
 FT DOMAIN 35 45 POLY-SER.  
 FT DOMAIN 56 59 POLY-THR.  
 FT DOMAIN 103 106 POLY-ALA.  
 FT DOMAIN 739 745 POLY-PRO.  
 FT DISULFID 273 327 BY SIMILARITY.  
 FT DISULFID 361 375 BY SIMILARITY.  
 FT DISULFID 369 386 BY SIMILARITY.  
 FT DISULFID 388 397 BY SIMILARITY.  
 FT CARBOHYD 33 33 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 34 34 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 362 362 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARSPLIC 1 108 MISSING (IN ISOFORM NRG2-BETA).  
 FT VARSPLIC 220 222 PLV -> FFF (IN ISOFORM NTAK-ALPHA2-1P).







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2003, 04:37:08 ; Search time 33.3512 Seconds  
(without alignments)  
3700.688 Million cell updates/sec

Title: US-09-775-181-4  
Perfect score: 3199  
Sequence: 1 MGAWAYPLLLCLLALQLGLG.....YMTAVGMSLVSYDCLATIFQ 599

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	339	10.6	669	5 Q9VR40	Q9vr40 drosophila
2	238.5	7.5	176	5 Q9WLP4	Q9wlp4 drosophila
3	181.5	5.7	594	5 O45500	O45500 caenorhabdi
4	149	4.7	1677	5 Q9VKA3	Q9vka3 drosophila
5	134.5	4.2	1305	5 Q9BML5	Q9bml5 drosophila
6	132.5	4.1	868	13 Q73636	Q73636 fugu rubrip
7	130.5	4.1	783	5 Q8SSS5	Q8sss5 dictyosteli
8	130	4.1	1305	5 Q9VPS7	Q9vps7 drosophila
9	127.5	4.0	840	11 Q9Z0R8	Q9z0r8 rattus norv
10	125	3.9	977	13 Q9PWE1	Q9pwe1 ictalurus p
11	124.5	3.9	528	5 Q96954	Q96954 geodia cydo
12	120.5	3.8	879	11 Q9QYS2	Q9qys2 mus musculu
13	120	3.8	738	5 Q9V4U3	Q9v4u3 drosophila
14	120	3.8	870	5 Q9N4T8	Q9n4t8 caenorhabdi
15	119.5	3.7	877	4 Q8TBH9	Q8tbh9 homo sapien
16	117	3.7	264	5 Q9WLP3	Q9wlp3 drosophila

## ALIGNMENTS

## RESULT 1

Q9VR40  
ID Q9VR40 PRELIMINARY; PRT; 669 AA.  
AC Q9VR40;  
DT 01-MAY-2000 (TREMREL. 13, Created)  
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)  
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)  
DE CGL1923 protein.  
GN CGL1923.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blasej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,





[illegible]

```
QY 332 LNNSECMPIKGLFVLGAYECI--CKAGFYHPGVLPVNNF----- 369
DB 467 --NGYAVSFANGDPVASYELYNWKKSGSSEVVPVGYDASLPQGEFRIFRDTIWD 524
QY 370 --RRGPDQIHSGS-----TKDVSEE-----AYVCLPCREG-----CPCADD----- 405
DB 525 GRKQVPVSVCSDCSQGTQKRVLKQKPKICCYCQVQCEGEISNVTDSPICPLCDDFWPN 584
QY 406 ---SPCFVQEDKYLR-----LATISFQGLCMLLDFVSMVLVYHFRKAKSIRASGLIL 454
DB 585 PERNACFPKPVFEELSGNEVLGILAVSFVGGACLA--ITAAVFFHRTSPIVRANNSL 642
QY 455 LETILFSGLLLYFPVVILYFEPSTCILLRWALLGFATVYGTVTLKLRHVKVFLSRT 514
DB 643 SFLLLSLTLCLFCLSTFTFGAPSHLSCLMLRHTAFGFTFVLCISCVLGTVVVLMFRAF-- 701
QY 515 AQRIPTWGTGRVW-----RMLAV-----ILLVWFELIGWTSVCONLEKQISLIG 560
DB 702 -----LPGSNVWKWFGPPQQRMTVVFTFSIQVLICIVLWLVVNPFPV--RNL----- 746
QY 561 QGKTSDLHIFNMCLDRWDYMTAVGMWSLVSYDGL 595
DB 747 --TTYKERIILECALG-----SSVGEFVAVLGYGL 774
RESULT 7
Q8SS5 PRELIMINARY; PRT; 783 AA.
AC Q8SS5;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE METABOTROPIC GABA-B receptor subtype 2. 6/101.
OS Dictyostelium discoideum (Slime mold). Dictyostelium.
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunngal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC115032; AAL93018.1; -.
SQ SEQUENCE 783 AA; 87581 MW; 91F1EC8FB923390D7 CRC64;
Query Match 4.1%; Score 130.5; DB 5; Length 783;
Best Local Similarity 21.7%; Pred. No. 0.019;
Matches 43; Conservative 44; Mismatches 86; Indels 25; Gaps 7;
QY 409 FVQEDKYLRLATISFQGLCMLLDFVSMVLVYHFRKAKSTRASSGLILLETILFSGLLLYXP 468
DB 372 FIDYPSNLKYGTIVSGVCIPTCLVCMVLVYVFKRARIKSSSPAFLLILGCGCIIFAA 431
QY 469 VVILYFEPSTCILLRWALLGFATVYGTVTLKLRHVKV-----LSRTA-----QRIPY 520
DB 432 CILFAGSPNTQCSARILLSLGYTLFGLNLLVKNRWIWLLEDNPKLKRATINWKLIPW 491
QY 521 MTGGVRVWRMLAV--ILLVWFELIGWTSVCONLEKQISLIGOGKTSDDLIFNMCLDRWD 579
DB 492 -----VFALIDVIMLAIWQGLG-----NINAE--SRIGYDSLTYQYQYKNVCSDD-- 536
QY 580 YMTAVGMWSLVSYDGLTI 597
DB 537 -QGSTALYLLLVFHLGLV 553
RESULT 8
Q9VPS7 PRELIMINARY; PRT; 1305 AA.
AC Q9VPS7;
```

```
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE CG3022 protein.
GN GABA-B-R3 OR CG3022.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Foslter C., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jajuli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Liu X., Mattei Y., McIntosh A.A., Li J., Li Z., Liang Y., Lin X.,
RA Munkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhou X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003588; AAF51465.2; -.
DR FlyBase; FBgn0031275; GABA-B-R3.
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR InterPro; IPR000005; HTHArac.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMRG.
DR PROSITE; PS50259; G_PROTEIN_RECEPTOR_F3_4; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; UNKNOWN_1.
SQ SEQUENCE 1305 AA; 143588 MW; A9C3C85307650450 CRC64;
Query Match 4.1%; Score 130; DB 5; Length 1305;
Best Local Similarity 19.2%; Pred. No. 0.043;
Matches 87; Conservative 68; Mismatches 164; Indels 134; Gaps 20;
QY 181 DLSLAP-APQVFLQATRESRIILLQDLSSAPH-----LANATLETFWHLGRKWRP 232
DB 383 ESMGAPWPDQRTACSNHELQLAVENTHNSIVGNVNNVSYGLNNHMFQURKQSA 442
QY 233 HLHRRGPNQGRGLGHSW-----RRKDLGLGDKSHFKWSPPYLECGNSY 277
```

```

Db 443 QFH-----GQDFGSGYGSRIISIAATQSDRRRRRRGGVGTSGGHLFPFAISQAPQY 496
QY 278 KPGWLVTLSAAYGLQPNLVPFRGVNVDINLQKVIDQCSSDGHGFSCTHKCHLNNSC 337
Db 497 DAYWATAIAL-----RAAEH-----RRNEQ 519
QY 338 MPITKGLGFGVLG--AYECICKAGFYH--PGVLPVNNFRRRGPDQHGIST-----KDVSEE 388
Db 520 SKLDGFDYTRSDMAWEFLQOMKHLHFGSGVSP--SGPDR--VGTTFAYQIQRLLEP 575
QY 389 AYVCLPREG-----CPFC-----ADDSPCFVQEDKYRLAIISF 423
Db 576 VALYYPATDALDFRCPCRPVKWHSQVPIAKRVFKLRVATIAPL-----AFYTIATLSS 630
QY 424 QGLCMLLDFVSMVYVHFRKASIRASGLILLETILFGSLLYFPVYLYFEPST-----478
Db 631 VGIATAIAELAFNL--HFRKKAIRKLSPLSNITAVGCFIVYATVILGLDHSPLPSAE 688
QY 479 --FRCILLRWARLL--GEATVYGTVTCLKHRLVKVFLSRTAQ-----RIPVMTGG 524
Db 689 DGFATVCTARVYLLSAGSLAFSGMFAKYRVHRIF--TRTGSVFVKMDQLQILLLVGG 747
QY 525 RVMRMLAVILLVFWPLIGWTSSVCQNKLEQIS 557
Db 748 ---LLILDALLVTLWVVTDPMERHLNLTLS 777

```

```

RESULT 9
Q3Z0R8
ID Q9Z0R8 PRELIMINARY; PRT; 840 AA.
AC Q9Z0R8;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative taste receptor TR1 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=99159821; PubMed=10052456;
RA Hoon M.A., Adler E., Lindemeier J., Battley J.F., Ryba N.J.,
RA Zuker C.S.;
RA "Putative mammalian taste receptors: a class of taste-specific GPCRs
with distinct topographic selectivity.";
Cell 96:541-551(1999).
DR EMBL; AF127389; AAD18069.1; -.
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS50259; G_PROTEIN_RECEP_F3_4; 2.
KW Receptor.
FT NON_TER.
SQ SEQUENCE 840 AA; 93496 MW; 1FCFB7EFC6B845DB CRC64;

Query Match 4.0%; Score 127.5; DB 11; Length 840;
Best Local Similarity 22.2%; Pred. No. 0.038;
Matches 60; Conservative 48; Mismatches 89; Indels 73; Gaps 14;

QY 306 VDINLQK-----VDIQCSSDGFSGTHKCHLNNSCEMPIKGLGVFGAYECI-CK 355
Db 475 LDINKTKIQWGHKNQVPVSVCTTD-CLAGHRRVYVVGSHHC-----CFECVPCPE 522
QY 356 AGEFYHPGVLPVNNFRRRGPDQHGISTKDVSEAYVCLPC--REGCPFCADDSPCFVQED 413
Db 523 ACTF-----LN-----MSELHICQPCGTEWAP--KESTTCFRTV 556
QY 414 KYL-----RLAIIISFQGLCMLLDF--VSMVYVYHFRKASIRASG-----LILLETILFGS 462

```

```

Db 557 EFLAWHEPISLVLIAANTLLLLLVGTAGLFAWHEH-TPVVRSGAGRLCFLMLGSLVAGS 615
QY 463 LLLYFPVILYEPSTFRCILLRWARLLGFAFYGTVTLKHLRVLKVLRSRTAQRIPLYMT 522
Db 616 CSFY-----SFFGEPTVPACLLRQPLFSGLFAIFLSCLTIRSFQLVIRFKFSTKVPFTYRT 671
QY 523 -----GGRVMMLAVILLVVFVFLGWT 545
Db 672 WAQNHGAGLFVIVSVSTVHLLICLTWLVMWT 701

RESULT 10
Q9PWEL PRELIMINARY; PRT; 977 AA.
ID Q9PWEL;
AC Q9PWEL;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Metabotropic glutamate receptor (Fragment).
GN GLUR.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ictalurus.
OX NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RA Hellmich H.L., Micci A.M., Sanchez X., Christensen B.N.;
RT "Molecular cloning, functional expression and localization of a novel
metabotropic glutamate receptor linked to calcium mobilization from
the catfish retina.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF076473; AADA7893.1; -.
DR InterPro; IPR001828; ANF_receptor.
DR InterPro; IPR000337; GPCR_Mgr.
DR Pfam; PF00003; 7tm_3; 1.
DR Pfam; PF01094; ANF_receptor; 1.
DR PRINTS; PR00248; GPCRMR.
DR PROSITE; PS00979; G_PROTEIN_RECEP_F3_1; 1.
DR PROSITE; PS00981; G_PROTEIN_RECEP_F3_3; 1.
DR PROSITE; PS50259; G_PROTEIN_RECEP_F3_4; 1.
KW Receptor.
FT NON_TER.
SQ SEQUENCE 977 AA; 108759 MW; BD06A67E240751CE CRC64;

Query Match 3.9%; Score 125; DB 13; Length 977;
Best Local Similarity 25.4%; Pred. No. 0.078;
Matches 45; Conservative 30; Mismatches 74; Indels 28; Gaps 7;

QY 392 CLPCR-----EGCPE-----CADDSPC-----FVQEDKYRLAI--SFQGLCMLLD 431
Db 572 CEPDGYQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 631
QY 432 FVSMVYV-HFRKASIRASGLILLETILFGSLLYFPVYLYFEPSTFRCILLRWARLL 490
Db 632 TLSVVIVFVRENDPTIVRASGRSLYVLLTGTLFIYITFLMIAEPNTVVCALRLLGL 691
QY 491 GFATVYGTVTCLKHRLVKV-----LSRTAQRIPLYMTGGVVRMLAV-----ILLVWF 540
Db 692 GMCITYSAMLTKTRIYRIFRQGGKSVTAPKFIPTSVLTILVSPQVIGVFIWF 748

RESULT 11
O96954 PRELIMINARY; PRT; 528 AA.
ID O96954;
AC O96954;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Metabotropic glutamate GABA-like receptor.
GN MGRL.

```











GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2003, 00:11:08 ; Search time 11.5573 Seconds  
(without alignments)  
1524.948 Million cell updates/sec

Title: US-09-775-181-4

Perfect score: 3199

Sequence: 1 MGAMAYPLLLCILLQAQLGLG.....YMTAVGMNLSVYDGLTIFQ 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*

3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133.5	4.2	1079	1 US-08-485-588-8	Sequence 8, Appli
2	133.5	4.2	1079	1 US-08-484-563-8	Sequence 8, Appli
3	133.5	4.2	1079	2 US-08-480-751-8	Sequence 8, Appli
4	133.5	4.2	1079	2 US-08-943-986-8	Sequence 8, Appli
5	133.5	4.2	1079	3 US-08-353-784-8	Sequence 8, Appli
6	133.5	4.2	1079	3 US-08-484-719B-8	Sequence 8, Appli
7	133.5	4.2	1079	4 US-08-484-159-8	Sequence 8, Appli
8	132	4.1	1027	4 US-09-162-021B-2	Sequence 2, Appli
9	127	4.0	1078	1 US-08-485-588-7	Sequence 7, Appli
10	127	4.0	1078	1 US-08-484-565-7	Sequence 7, Appli
11	127	4.0	1078	2 US-08-480-751-7	Sequence 7, Appli
12	127	4.0	1078	2 US-08-943-986-7	Sequence 7, Appli
13	127	4.0	1078	3 US-08-353-784-7	Sequence 7, Appli
14	127	4.0	1078	3 US-08-484-719B-7	Sequence 7, Appli
15	127	4.0	1078	4 US-08-484-159-7	Sequence 7, Appli
16	123	3.8	1085	1 US-08-485-588-5	Sequence 5, Appli
17	123	3.8	1085	1 US-08-484-565-5	Sequence 5, Appli
18	123	3.8	1085	2 US-08-480-751-5	Sequence 5, Appli
19	123	3.8	1085	2 US-08-943-986-5	Sequence 5, Appli
20	123	3.8	1085	3 US-08-353-784-5	Sequence 5, Appli
21	123	3.8	1085	3 US-08-484-719B-5	Sequence 5, Appli
22	123	3.8	1085	4 US-08-484-159-5	Sequence 5, Appli
23	121.5	3.8	1059	4 US-09-134-513-2	Sequence 2, Appli
24	119.5	3.7	1088	1 US-08-485-588-6	Sequence 6, Appli
25	119.5	3.7	1088	1 US-08-484-565-6	Sequence 6, Appli
26	119.5	3.7	1088	2 US-08-480-751-6	Sequence 6, Appli
27	119.5	3.7	1088	2 US-08-943-986-6	Sequence 6, Appli

28	119.5	3.7	1088	3 US-08-353-784-6	Sequence 6, Appli
29	119.5	3.7	1088	3 US-08-484-719B-6	Sequence 6, Appli
30	119.5	3.7	1088	4 US-08-484-159-6	Sequence 6, Appli
31	118.5	3.7	1058	2 US-08-687-289A-5	Sequence 5, Appli
32	118	3.7	872	3 US-08-337-797A-2	Sequence 2, Appli
33	118	3.7	872	3 US-09-258-523-2	Sequence 2, Appli
34	116	3.6	906	1 US-08-486-270-2	Sequence 2, Appli
35	116	3.6	906	3 US-08-367-264-2	Sequence 2, Appli
36	116	3.6	906	4 US-09-153-757-2	Sequence 2, Appli
37	116	3.6	1056	2 US-08-687-289A-8	Sequence 8, Appli
38	116	3.6	1194	4 US-08-538-526-1	Sequence 1, Appli
39	115.5	3.6	879	4 US-08-794-158-2	Sequence 2, Appli
40	115	3.6	906	5 PCT-US91-09422-17	Sequence 17, Appli
41	115	3.6	1199	1 US-08-041-538-2	Sequence 2, Appli
42	115	3.6	1199	1 US-08-463-642-2	Sequence 2, Appli
43	115	3.6	1199	1 US-08-455-602-2	Sequence 2, Appli
44	115	3.6	1199	2 US-08-465-157-2	Sequence 2, Appli
45	115	3.6	1199	5 PCT-US91-09422-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-08-485-588-8

; Sequence 8, Application US/08485588

; Patent No. 5688938

; GENERAL INFORMATION:

; APPLICANT: Edward M. Brown

; APPLICANT: Steven C. Hebert

; APPLICANT: Forrest H. Fuller

; APPLICANT: James E. Garrett, Jr.

; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE

; MOLECULES

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lyon & Lyon

; STREET: First Interstate World Center

; STREET: Suite 4700

; STREET: 633 West Fifth Street

; CITY: Los Angeles

; STATE: California

; COUNTRY: USA

; ZIP: 90071

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: FASTSEQ

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/485,588

; FILING DATE: 7 June, 1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; PRIOR APPLICATION DATA: including application

; PRIOR APPLICATION DATA: described below: 9

; APPLICATION NUMBER: 08/353,784

; FILING DATE: 9 December, 1994

; APPLICATION NUMBER: PCT/US94/12117

; FILING DATE: 21 October, 1994

; APPLICATION NUMBER: U.S. 08/292,827

; FILING DATE: 23 August, 1994

; APPLICATION NUMBER: U.S. 08/141,248

; FILING DATE: 22 October, 1993

; APPLICATION NUMBER: U.S. 08/009,389

; FILING DATE: 23 February, 1993

; APPLICATION NUMBER: U.S. 08/017,127

; FILING DATE: 12 February, 1993

; APPLICATION NUMBER: U.S. 07/934,161

; FILING DATE: 21 August, 1992

; APPLICATION NUMBER: U.S. 07/834,044

; FILING DATE: 11 February, 1992

; APPLICATION NUMBER: U.S. 07/749,451



Db 668 FIGEPQDWTCRLRQPAFGISFVLCISCLVKNRVLVFEAKIPTSFHRKMWGLNLQFL 727  
QY 532 V-----ILLVVFWMFLIGWTSVQNLKQISLIGQKTSDBLIFNMCLIDRWDMYAV 584  
Db 728 VFLCTFMQILICIWIYTAPPSSY-RNHELE-----DEIIFITC---HEGSLMAL 773  
QY 585 GMWSLVSY 592  
Db 774 G--SLIGY 779

RESULT 3

US-08-480-751-8  
; Sequence 8, Application US/08480751  
; Patent No. 5858684  
; GENERAL INFORMATION:  
; APPLICANT: Edward F. Nemeth  
; APPLICANT: Edward M. Brown  
; APPLICANT: Steven C. Hebert  
; APPLICANT: Forrest H. Fuller  
; APPLICANT: James E. Garrett, Jr.  
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
; TITLE OF INVENTION: MOLECULES  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: First Interstate World Center  
; STREET: Suite 4700  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: FASTSEQ  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,751  
; FILING DATE: 7 June, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION DATA: including application  
; PRIOR APPLICATION DATA: described below: 9  
; APPLICATION NUMBER: 08/353,784  
; FILING DATE: 9 December, 1994  
; APPLICATION NUMBER: PCT/US/94/12117  
; FILING DATE: 21 October, 1994  
; APPLICATION NUMBER: U.S. 08/292,827  
; FILING DATE: 23 August, 1994  
; APPLICATION NUMBER: U.S. 08/141,248  
; FILING DATE: 22 October, 1993  
; APPLICATION NUMBER: U.S. 08/009,389  
; FILING DATE: 23 February, 1993  
; APPLICATION NUMBER: U.S. 08/017,127  
; FILING DATE: 12 February, 1993  
; APPLICATION NUMBER: U.S. 07/934,161  
; FILING DATE: 21 August, 1992  
; APPLICATION NUMBER: U.S. 07/834,044  
; FILING DATE: 11 February, 1992  
; APPLICATION NUMBER: U.S. 07/749,451  
; FILING DATE: 23 August, 1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Heber, Sheldon O.  
; REGISTRATION NUMBER: 38,179  
; REFERENCE/DOCKET NUMBER: 213/004  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1079 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-480-751-8  
Query Match 4.2%; Score 133.5; DB 2; Length 1079;  
Best Local Similarity 21.7%; Pred. No. 0.00056;  
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;  
QY 235 HRRGPNQGRGLGHSRRKDGKSHKSPPLYCENGSKYKFGWLVTLSAIYGLQP 294  
Db 377 HEEG---GNRLNLSSTAFRPLCTGDENINSVETPYMDYEHLSRISYNYVLAVIYIAHALQD 433  
QY 295 --NLVPEFRGMKVDINLQKVIDIDOCSSDGWFSGTHKCHLNN-----ECMPIKG 342  
Db 434 IYTCPLG-RGLF---TNGSCADIKV--EAWQVLKHLRLHNLFTNMGEQVTFDECGDIVG 487  
QY 343 LGFVL-----GAYECICKAG---FYHPCVLFPVNNFRRRGPDPQHS---- 379  
Db 488 NYSIINWHLSPEDGSIVFKEGVYNYAKKGERLFINEEKLWWSGFSREVFPFNSCRDQ 547  
QY 380 -GSTKDVSEE-----AYVCLPREG-----CPFCADD-----SPCFVQEDKYLR 417  
Db 548 AGTRKGIIEGEPTCCFECVCPDGEYSGETDASACDKCPDFWSNENHTSCIAKEIEFLA 607  
QY 418 -----LAISFOGLCMLLDFVSMVLVYHFRKAKSTRASGLILLETILFGSLLLFPVVI 471  
Db 608 WTEPGIALTLFAVLGIFLTAFLVGFKFRWTPIVKATNRELSYLLLSLCCFSSSLF 667  
QY 472 LYFEPSTFRCLLRWARLLGFATVYGTVLKHLRVKLVLSRTAQRIPTYMTGGRVMRLA 531  
Db 668 FIGEPQDWTCRLRQPAFGISFVLCISCLVKNRVLVFEAKIPTSFHRKMWGLNLQFL 727  
QY 532 V-----ILLVVFWMFLIGWTSVQNLKQISLIGQKTSDBLIFNMCLIDRWDMYAV 584  
Db 728 VFLCTFMQILICIWIYTAPPSSY-RNHELE-----DEIIFITC---HEGSLMAL 773  
QY 585 GMWSLVSY 592  
Db 774 G--SLIGY 779

RESULT 4

US-08-943-986-8  
; Sequence 8, Application US/08943986  
; Patent No. 5962314  
; GENERAL INFORMATION:  
; APPLICANT: Edward M. Brown  
; APPLICANT: Steven C. Hebert  
; APPLICANT: James E. Garrett, Jr.  
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
; TITLE OF INVENTION: MOLECULES  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: First Interstate World Center  
; STREET: Suite 4700  
; STREET: 633 West Fifth Street  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: FASTSEQ  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,986  
; FILING DATE: 03-OCT-1997  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/484,565  
FILING DATE: 7-June-1995  
APPLICATION NUMBER: 08/353,784  
FILING DATE: 9 December, 1994  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994  
APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451  
FILING DATE: 23 August, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 213/006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1079 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-943-986-8

Query Match 4.2% Score 133.5; DB 2: Length 1079;  
Best Local Similarity 21.7% Pred. No. 0.00056;  
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;  
QY 235 HRRGNQGRGLGHSWRKDKGLGGKSHFKWSPYLECENGSKYKPGWLVTLSAIYGLQP 294  
DB 377 HEEG---GNRLNSSTAFRPLCTGDNINSVETPYMDYHLRISYNYLVAVYSAHAHQD 433  
QY 295 --NLVPERGVKMDINLQKVDIDOCSSDGFSGTHKCHLANS-----ECMPIKG 342  
434 IYTCPLG-RGLF---TNGSCADIKV--EAMOVKLHLRHLNFTNMGQVTFDECGDLVG 487  
343 LGFVL-----GAYECICKAG---FYHPGVLPVNNFRRRGPDQHIS----- 379  
488 NYSIINHLSPEDGSIKVEGYNNYAKGGERLFINEEKILWSGFSREVFPSCNCRDQ 547  
QY 380 -GSTKDVSEE----AYVCLPREG-----CPFCADD-----SPCFVQEDKYL 417  
548 AGTRKGIEGTEPTCCFECVECPDGEYSGETDASACDKPCDDFWSNENHTSIAKEIFLA 607  
QY 418 -----LALISPGQLMLLDFVSMVLVYVHERKAKSIRASGLILLETILFGSLLLYFPVVI 471  
608 WTEPPGIALTFVAVLGIFELTAFVLGVFTKFRNTPIVKATNRELVSLLLSLCCFSSSLF 667  
QY 472 LYFEPSTFCILLRWARLLGFATVYGTVTLKHLRVKLVFLSRTAQRIPYMTGGRVWRMLA 531  
668 FIGEPQDWTCRLRQAFGISVFLCISCLVKTNRVLLVFEAKIPTSPFRKWGNLQFLL 727  
QY 532 V-----ILLVWFVLGWTSSVCQNLKQISLIGQCKTSDHLIFNMCCLIDRWYMTAV 584  
728 VFLCTFMQILICIIIMLYTAPPSY--RNHELE-----DEIIFITC---HEGSLMAL 773  
QY 585 GNMVSLVSY 592  
DB 774 G---SLUGY 779

RESULT 5  
US-08-353-784-8  
Sequence 8, Application US/08353784  
Patent No. 6011068  
GENERAL INFORMATION:  
APPLICANT: Edward F. Nemeth, Edward M.  
APPLICANT: Brown, Steven C. Hebert,  
APPLICANT: Bradford C. Van Wagenen, Manuel  
APPLICANT: F. Balandrin, Forrest H. Fuller,  
APPLICANT: Eric G. DelMar, and Scott T. Moe  
TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
MOLECULES  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: First Interstate World Center  
STREET: Suite 4700  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,784  
FILING DATE: 9 December, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA: including application  
PRIOR APPLICATION DATA: described below: 8  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994  
APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451  
FILING DATE: 23 August, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 209/069  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1079 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-353-784-8

Query Match 4.2% Score 133.5; DB 3: Length 1079;  
Best Local Similarity 21.7% Pred. No. 0.00056;  
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;  
QY 235 HRRGNQGRGLGHSWRKDKGLGGKSHFKWSPYLECENGSKYKPGWLVTLSAIYGLQP 294  
DB 377 HEEG---GNRLNSSTAFRPLCTGDNINSVETPYMDYHLRISYNYLVAVYSAHAHQD 433



NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: First Interstate World Center  
STREET: Suite 4700  
STREET: 633 West Fifth Street  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: FASTSEQ  
CURRENT APPLICATION DATA:  
APPLICANT: Edward M. Brown  
FILING DATE: 7 June, 1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
PRIORITY APPLICATION DATA: including application  
PRIORITY APPLICATION DATA: described below: 9  
APPLICATION NUMBER: 08/353,784  
FILING DATE: 9 December, 1994  
APPLICATION NUMBER: PCT/US/94/12117  
FILING DATE: 21 October, 1994  
APPLICATION NUMBER: U.S. 08/292,827  
FILING DATE: 23 August, 1994  
APPLICATION NUMBER: U.S. 08/141,248  
FILING DATE: 22 October, 1993  
APPLICATION NUMBER: U.S. 08/009,389  
FILING DATE: 23 February, 1993  
APPLICATION NUMBER: U.S. 08/017,127  
FILING DATE: 12 February, 1993  
APPLICATION NUMBER: U.S. 07/934,161  
FILING DATE: 21 August, 1992  
APPLICATION NUMBER: U.S. 07/834,044  
FILING DATE: 11 February, 1992  
APPLICATION NUMBER: U.S. 07/749,451  
FILING DATE: 23 August, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Heber, Sheldon O.  
REGISTRATION NUMBER: 38,179  
REFERENCE/DOCKET NUMBER: 214/101  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 935-0440  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1079 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-484-159-8

Query Match 4.2%; Score 133.5; DB 4; Length 1079;  
Best Local Similarity 21.7%; Pred. No. 0.00056;  
Matches 93; Conservative 58; Mismatches 182; Indels 95; Gaps 18;  
QY 235 HRRGNPGRLGHWSRRKDGGLGGKSHFKWSPPYLECEGNSYKPGWLVTLSAIIYGLQP 294  
DB 377 HEEG---GNRLNLSSTAFRLCTGDNINSVETPYMDYEHRLRSYNYVLAIVSTAHLQD 433  
QY 295 ---NLVPEFRGVMKVDINLQKVDICSSDGFSGTHKCHLNN-----ECMPTKG 342  
DB 434 IYTCPLPG-RGLF---TNGSCADIKV--EAWQVLKHLRLNFTNNMGQVTFDECGDLVG 487  
QY 343 LGFVL-----GAYECICKAG---FYHPCVLPVNNFRRGPDQHIS----- 379  
DB 488 NVSIIINWHLSPEDGSIVPEKVEGYNVYAKGERLFINKEILWGSFSEVPFNSCRDQ 547  
QY 380 -GSTKDVSEE---AYVCLPCREG-----CPFCADD-----SPCFVQEDKYL 417

Db 548 AGTRKGIIEGEPTCCFCEVCEPCDGEYSGETDASACDKCPDDFWSNENHTSCIAKEIEFLA 607  
QY 418 -----LAIISFOGLCMLLDVFSMLVYVYHFRKAKSIRASGLILLETILFGSLLLYFPVVI 471  
DB 608 WTEPFGIALTLFAVLGIFLTAFLVGLGVFIKFRNTPTVKATNRELSYLLLFSLCCCFSSSLF 667  
QY 472 LYFEPSTFRCILLRWARLLGFATVGTVTLKLHRLVKVFLSRTAQRIPIYMTGGRVMRLA 531  
DB 668 FIEPQDWTCLRPAPFGISVLCISLIVKTNRVLLVFEAKIPTSFHRKWKWGLNLQELL 727  
QY 532 V-----ILLVWFVFLICWTSSVCNLEKQISLIGOKTSDHLIFNMCLIDRMDYMTAV 584  
DB 728 VFLCTFMOLLCIIWLTYAPFSSY-RNHELE-----DEIFITC---HEGSLMAL 773  
QY 585 GWSLVSY 592  
DB 774 G--SLIGY 779  
RESULT 8  
US-09-162-021B-2  
Sequence 2, Application US/09162021B  
Patent No. 6337391  
GENERAL INFORMATION:  
APPLICANT: H. William Harris  
APPLICANT: Edward M. Brown  
APPLICANT: Steven C. Hebert  
TITLE OF INVENTION: Polycation-Sensing Receptor in Aquatic  
FILE REFERENCE: 2856.1001-007  
CURRENT APPLICATION NUMBER: US/09/162.021B  
CURRENT FILING DATE: 1998-09-28  
PRIOR APPLICATION NUMBER: PCT/US97/05031  
PRIOR FILING DATE: 1997-03-27  
PRIOR APPLICATION NUMBER: 08/622,738  
PRIOR FILING DATE: 1996-03-27  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 1027  
TYPE: PRT  
ORGANISM: squalus acanthias  
US-09-162-021B-2

Query Match 4.1%; Score 132; DB 4; Length 1027;  
Best Local Similarity 19.5%; Pred. No. 0.00073;  
Matches 88; Conservative 58; Mismatches 166; Indels 140; Gaps 17;  
QY 239 PNOGPRGLGHSWR---RKDGL---GGDKSHFKWSPPYLECEGNSYKPGWLVTLSAIIY 290  
DB 378 PSHGPAAGDGSGKAGNSRRRTALRHPCGTGEENITSVETPYLDYTHLRISYNYVAVYIAH 437  
QY 291 GLQPNLVPEFRGVMKVDINLQKVDIDOCSS-----DGWFSGTHKCHLNN 334  
DB 438 ALQ-----DIHSCKPGTGIFANGSCADIKKVEAWQVLNHLHLKF 477  
QY 335 SEC-----PIRG-----LGFVLGAYECICK----- 355  
DB 478 TNSMGQVDFDODGLKGNVTIINWQLSAEDESVLFEHGVNAYAKPSDRNLNNEKKIL 537  
QY 356 -AGFYHPCVLPVNNFRRGPDQHISGSKDVSEE---AYVCLPCREG-----CP 400  
DB 538 WSGF--SKVVPFNSCSR---DCVPGTGRKGIIEGPTCCFECMACAGEFSDENDASACT 591  
QY 401 FCADD-----SPCFVQEDKYL-----LAIISFOGLCMLLDVFSMLVYVYHFRKAKSI 447  
DB 592 KCPNDFWSNENHTSCIAKEIYLSWTBPFGLIATIFAVLGILITFSVLGVFIKERNPTIV 651  
QY 448 RASGLIILETLFGSLLLYFPVVIYFEPSTFRCILLRWARLLGFATVGTVTLKLHRLV 507  
DB 652 KATNRELSYLLLFSLICCFSSSLIFIGEPDWTCLRPAPFGISVLCISLIVKTNRVL 711



QY 508 KVFSLRTAQIRPYMTGGRVWRMLAVLLV-----VFVFLIGWTSSVCQNKLEKQISLIG 560  
 Db 712 LVFEAKIPTSLHRKWKVGLNLQFLVFLCIVLQVTCIIWLYTAPPSSY-RNHELE----- 765  
 QY 561 QGKTSDFHFNCLDRWDYMTAVGWSLVSY 592  
 Db 766 -----DEVIFITC-----DEGSLMALGFLIGY 787

RESULT 9  
 US-08-485-588-7  
 ; Sequence 7, Application US/08485588  
 ; Patent No. 5688938  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edward M. Brown  
 ; APPLICANT: Steven C. Hebert  
 ; APPLICANT: Forrest H. Fuller  
 ; APPLICANT: James E. Garrett, Jr.  
 ; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: First Interstate World Center  
 ; STREET: Suite 4700  
 ; STREET: 633 West Fifth Street  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 90071

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: FASTSEQ  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/485,588  
 ; FILING DATE: 7 June, 1995  
 ; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 ; INCLUDING APPLICATION  
 ; PRIOR APPLICATION DATA: described below: 9  
 ; APPLICATION NUMBER: 08/353,784  
 ; FILING DATE: 9 December, 1994  
 ; APPLICATION NUMBER: PCT/US/94/12117  
 ; FILING DATE: 21 October, 1994  
 ; APPLICATION NUMBER: U.S. 08/292,827  
 ; FILING DATE: 23 August, 1994  
 ; APPLICATION NUMBER: U.S. 08/141,248  
 ; FILING DATE: 22 October, 1993  
 ; APPLICATION NUMBER: U.S. 08/009,389  
 ; FILING DATE: 23 February, 1993  
 ; APPLICATION NUMBER: U.S. 08/017,127  
 ; FILING DATE: 12 February, 1993  
 ; APPLICATION NUMBER: U.S. 07/934,161  
 ; FILING DATE: 21 August, 1992  
 ; APPLICATION NUMBER: U.S. 07/834,044  
 ; FILING DATE: 11 February, 1992  
 ; APPLICATION NUMBER: U.S. 07/749,451  
 ; FILING DATE: 23 August, 1991

ATTORNEY/AGENT INFORMATION:  
 ; NAME: Heber, Sheldon O.  
 ; REGISTRATION NUMBER: 38,179  
 ; REFERENCE/DOCKET NUMBER: 213/005  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (213) 489-1600  
 ; TELEFAX: (213) 955-0440  
 ; TELE: 67-3510

INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1078 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear

MOLECULE TYPE: protein  
 US-08-485-588-7  
 Query Match 4.0%; Score 127; DB 1; Length 1078;  
 Best Local Similarity 22.5%; Fred. No. 0.0025;  
 Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;  
 QY 311 QKVIDIQSSDGFSGTHKC-----HLNNE-CMPIKGLGLVGLGAYECICKAG---FYHPGV 363  
 Db 475 EQVTEDEC---GDIVGNYSIINWHLSPEDGSIVFVEGVY---YVNYAKKGERLEINEEK 527  
 QY 364 LPVNNFRRRGPQH-----ISGSTDVSEB-----AYVCLPCREG-----CPFCAD 404  
 Db 528 ILWSGFSREVPEPSCNRDCLAGTRKGIIEGTEPTCCFCEVCEPCDGEYSDETDASACNKKCPD 587  
 QY 405 D-----SPCFVQEDKYLK-----LAIISQGLCMLLDVFSMLVYVYHFRKAKSIRASG 451  
 Db 588 DFWSNENHTSCIAKEIEFLSWTEPEFGIALTLFAVLGIFLTAFLVGVFIKFRNTPIVKATN 647  
 QY 452 LILLETILFGLSLLYFPVVVILYFEPSTFRCLLRWARLIGFATVYGTVTLKLRVLKVF 511  
 Db 648 RELSYLLLFSLCCFSSSLFFIGEPODWTCLRLQPAFGISFVLCISCLVKTNRVLLVFE 707  
 QY 512 SRTAQIRPYMTGGRVWRMLAV-----ILLVVFVFLIGWTSSV-CQNKLEKQISLI----- 559  
 Db 708 AKIPTSFHRKWKVGLNLQFLVFLCTFMQIVICVILWYTAPPSSYRNQOELEDEIIFITCHE 767  
 QY 560 GQKTSDFHFNCLLI 575  
 Db 768 GSLMALGFLIGYTCLL 783

RESULT 10  
 US-08-484-565-7  
 ; Sequence 7, Application US/08484565  
 ; Patent No. 5763569  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edward M. Brown  
 ; APPLICANT: Steven C. Hebert  
 ; APPLICANT: James E. Garrett, Jr.  
 ; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: First Interstate World Center  
 ; STREET: Suite 4700  
 ; STREET: 633 West Fifth Street  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: FASTSEQ  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/484,565  
 ; FILING DATE: 7 June, 1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; INCLUDING APPLICATION  
 ; PRIOR APPLICATION DATA: described below: 9  
 ; APPLICATION NUMBER: 08/353,784  
 ; FILING DATE: 9 December, 1994  
 ; APPLICATION NUMBER: PCT/US/94/12117  
 ; FILING DATE: 21 October, 1994  
 ; APPLICATION NUMBER: U.S. 08/292,827  
 ; FILING DATE: 23 August, 1994  
 ; APPLICATION NUMBER: U.S. 08/141,248  
 ; FILING DATE: 22 October, 1993  
 ; APPLICATION NUMBER: U.S. 08/009,389  
 ; FILING DATE: 23 February, 1993  
 ; APPLICATION NUMBER: U.S. 08/017,127  
 ; FILING DATE: 12 February, 1993  
 ; APPLICATION NUMBER: U.S. 07/934,161  
 ; FILING DATE: 21 August, 1992  
 ; APPLICATION NUMBER: U.S. 07/834,044  
 ; FILING DATE: 11 February, 1992  
 ; APPLICATION NUMBER: U.S. 07/749,451  
 ; FILING DATE: 23 August, 1991

```

; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; FILING DATE: 12 February, 1993
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 213/006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-484-565-7

Query Match          4.0%; Score 127; DB 1; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0025;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVIDQSSDGNFSGTHKC-----HNNSE-CMPIKGLGFLVGLAYECICKAG---FYHPGV 363
Db 475 EQVTFDEC---GDLVGNYSIINHLSPEDEGSIYFKEVGY---YNNYAKKGERLFINEEK 527
QY 364 LPVNNFRRRGPDQH-----ISGSTDVSEE---AYVCLPCREG-----CPFCAD 404
Db 528 ILWSGFSREVPFNSCRDLAGTRKGIIEGPTCCFCEVCEPDGSEYSDTDASACNKC 587
QY 405 D-----SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVHFRKAKSIRASG 451
Db 588 DFWSENHTSCIAKEIEFLSWTEPFPGIATLFAVLGIFLFAVLGVFKFRNTPIVKATN 647
QY 452 LILLETILFGSLLLYFPVYILYFEPSTFCILLRWALLGFATYVGTVTLKLRVLRVFL 511
Db 648 RELSYLLLFSLCCFSSSLFFIGEPQDWTCLRQPAFGISFVLCISCLIVKTNRVLLVFE 707
QY 512 SRTAQRIPYMTGGRVMRLAV-----ILLVWFELIGWTSV-CQNLKQISLI----- 559
Db 708 AKIPTSFRKRWGMLNQLFLLVFLCTFMQIVICVIWLYTAPPSSYRNQOELEDEIIFITCHE 767

QY 560 GOGKTSDDLIFNMCL 575
Db 768 GSLMALGFLIGYTCLL 783

RESULT 11
US-08-480-751-7
; Sequence 7, Application US/08480751
; Patent No. 5858684
; GENERAL INFORMATION:
; APPLICANT: Edward F. Nemeth
; APPLICANT: Edward M. Brown
; APPLICANT: Steven C. Hebert
; APPLICANT: Forrest H. Fuller
; APPLICANT: James E. Garrett, Jr.
; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Lyon & Lyon
; STREET: First Interstate World Center
; STREET: Suite 4700
; STREET: 633 West Fifth Street
; CITY: Los Angeles

```

```

; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,751
; FILING DATE: 7 June, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 9
; APPLICATION NUMBER: 08/353,784
; FILING DATE: 9 December, 1994
; APPLICATION NUMBER: PCT/US/94/12117
; FILING DATE: 21 October, 1994
; APPLICATION NUMBER: U.S. 08/292,827
; FILING DATE: 23 August, 1994
; APPLICATION NUMBER: U.S. 08/141,248
; FILING DATE: 22 October, 1993
; APPLICATION NUMBER: U.S. 08/009,389
; FILING DATE: 23 February, 1993
; APPLICATION NUMBER: U.S. 08/017,127
; APPLICATION NUMBER: U.S. 07/934,161
; FILING DATE: 21 August, 1992
; APPLICATION NUMBER: U.S. 07/834,044
; FILING DATE: 11 February, 1992
; APPLICATION NUMBER: U.S. 07/749,451
; FILING DATE: 23 August, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Heber, Sheldon O.
; REGISTRATION NUMBER: 38,179
; REFERENCE/DOCKET NUMBER: 213/004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-751-7

```

```

Query Match          4.0%; Score 127; DB 2; Length 1078;
Best Local Similarity 22.5%; Pred. No. 0.0025;
Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;

QY 311 QKVIDQSSDGNFSGTHKC-----HNNSE-CMPIKGLGFLVGLAYECICKAG---FYHPGV 363
Db 475 EQVTFDEC---GDLVGNYSIINHLSPEDEGSIYFKEVGY---YNNYAKKGERLFINEEK 527
QY 364 LPVNNFRRRGPDQH-----ISGSTDVSEE---AYVCLPCREG-----CPFCAD 404
Db 528 ILWSGFSREVPFNSCRDLAGTRKGIIEGPTCCFCEVCEPDGSEYSDTDASACNKC 587
QY 405 D-----SPCFVQEDKYLR-----LAIISFQGLCMLLDFVSMVYVHFRKAKSIRASG 451
Db 588 DFWSENHTSCIAKEIEFLSWTEPFPGIATLFAVLGIFLFAVLGVFKFRNTPIVKATN 647
QY 452 LILLETILFGSLLLYFPVYILYFEPSTFCILLRWALLGFATYVGTVTLKLRVLRVFL 511
Db 648 RELSYLLLFSLCCFSSSLFFIGEPQDWTCLRQPAFGISFVLCISCLIVKTNRVLLVFE 707
QY 512 SRTAQRIPYMTGGRVMRLAV-----ILLVWFELIGWTSV-CQNLKQISLI----- 559
Db 708 AKIPTSFRKRWGMLNQLFLLVFLCTFMQIVICVIWLYTAPPSSYRNQOELEDEIIFITCHE 767

```

QY 560 GQKTSDDLIFNMCLI 575  
 Db 768 GSLMALGFLIGYTCLL 783

RESULT 12

US-08-943-986-7  
 ; Sequence 7, Application US/08943986  
 ; Patent No. 5962314  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edward M. Brown  
 ; APPLICANT: Steven C. Hebert  
 ; APPLICANT: James E. Garrett, Jr.  
 ; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
 ; TITLE OF INVENTION: MOLECULES  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: First Interstate World Center  
 ; STREET: Suite 4700  
 ; STREET: 633 West Fifth Street  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: FASTSEQ  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/943,986  
 ; FILING DATE: 03-OCT-1997  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/484,565  
 ; FILING DATE: 7-June-1995  
 ; APPLICATION NUMBER: 08/353,784  
 ; FILING DATE: 9 December, 1994  
 ; APPLICATION NUMBER: PCT/US/94/12117  
 ; FILING DATE: 21 October, 1994  
 ; APPLICATION NUMBER: U.S. 08/292,827  
 ; FILING DATE: 23 August, 1994  
 ; APPLICATION NUMBER: U.S. 08/141,248  
 ; FILING DATE: 22 October, 1993  
 ; APPLICATION NUMBER: U.S. 08/009,389  
 ; FILING DATE: 23 February, 1993  
 ; APPLICATION NUMBER: U.S. 08/017,127  
 ; FILING DATE: 12 February, 1993  
 ; APPLICATION NUMBER: U.S. 07/934,161  
 ; FILING DATE: 21 August, 1992  
 ; APPLICATION NUMBER: U.S. 07/834,044  
 ; FILING DATE: 11 February, 1992  
 ; APPLICATION NUMBER: U.S. 07/749,451  
 ; FILING DATE: 23 August, 1991  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Heber, Sheldon O.  
 ; REGISTRATION NUMBER: 38,179  
 ; REFERENCE/DOCKET NUMBER: 213/006  
 ; TELEPHONE: (213) 489-1600  
 ; TELEFAX: (213) 955-0440  
 ; TELEX: 67-3510  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1078 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-943-986-7

Query Match 4.0%; Score 127; DB 2; Length 1078;  
 Best Local Similarity 22.5%; Pred. No. 0.0025;

Matches 71; Conservative 48; Mismatches 139; Indels 58; Gaps 13;  
 QY 311 OKVDIDQSSDGSFSGTHKC---HLNNS-CMPDKGLGFLVAYECICKAG---FYHGV 363  
 Db 475 EQVTFDEC---GDLVGNYSIIINWHLSPEDGSIVFKEVGY---YNYAKKGERLFINEEK 527  
 QY 364 LPVNNFRRRGPDQH-----ISGSKDVSEE---AYVCLPREG-----CPFCAD 404  
 Db 528 ILWSGFSREVPPFNSCRDCLAGTRKGIIEGEPCCFCEVCEPDGEYSDETDASACNKCPD 587  
 QY 405 D-----SPCFVQEDKYLK-----LAISFOGLCMLDLDFVSMVYVYHFRKAKSIRASG 451  
 Db 588 DFWSNNHTSCIAKETEFLSWTEPFPGIALTLFAVLGIFLTAFLVGLVFIKFRNTPIVKATN 647  
 QY 452 LILLETILFGLSLLYFPVVILYPESTFRCILLRWARLLGFAVYGVTVTLKLRVLKRVFL 511  
 Db 648 RELSYLLLSLCCFSSSLFFIGEPODWTCRLRQPAFGISFVLCISLILVKTNRVLLVFE 707  
 QY 512 SRTAQRIPTYMTGGRVMRLAV-----ILLVVFVFLIGWTSSV-CQNLEKQISLI---- 559  
 Db 708 AKIPTSFRKRWGLNLQFLVFLCTFMQIVICVILWLYTAPPSSYRNOELEDIEIFITCHE 767  
 QY 560 GQKTSDDLIFNMCLI 575  
 Db 768 GSLMALGFLIGYTCLL 783

RESULT 13

US-08-353-784-7  
 ; Sequence 7, Application US/08353784  
 ; Patent No. 6011068  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edward F. Nemeth, Edward M.  
 ; APPLICANT: Brown, Steven C. Hebert, Manuel  
 ; APPLICANT: Bradford C. Van Wagenen, Manuel  
 ; APPLICANT: F. Balandrin, Forrest H. Fuller,  
 ; APPLICANT: Eric G. DelMar, and Scott T. Moe  
 ; TITLE OF INVENTION: CALCIUM RECEPTOR-ACTIVE  
 ; TITLE OF INVENTION: MOLECULES  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Lyon & Lyon  
 ; STREET: First Interstate World Center  
 ; STREET: Suite 4700  
 ; STREET: 633 West Fifth Street  
 ; CITY: Los Angeles  
 ; STATE: California  
 ; COUNTRY: USA  
 ; ZIP: 90071  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: FASTSEQ  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/353,784  
 ; FILING DATE: 9 December, 1994  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; PRIOR APPLICATION DATA: including application  
 ; PRIOR APPLICATION DATA: described below: 8  
 ; APPLICATION NUMBER: PCT/US/94/12117  
 ; FILING DATE: 21 October, 1994  
 ; APPLICATION NUMBER: U.S. 08/292,827  
 ; FILING DATE: 23 August, 1994  
 ; APPLICATION NUMBER: U.S. 08/141,248  
 ; FILING DATE: 22 October, 1993  
 ; APPLICATION NUMBER: U.S. 08/009,389  
 ; FILING DATE: 23 February, 1993  
 ; APPLICATION NUMBER: U.S. 08/017,127  
 ; FILING DATE: 12 February, 1993  
 ; APPLICATION NUMBER: U.S. 07/934,161  
 ; FILING DATE: 21 August, 1992







GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 13, 2003, 00:15:14 ; Search time 8.58545 Seconds  
(without alignments)  
1782.525 Million cell updates/sec

Title: US-09-775-181-4

Perfect score: 3199

Sequence: 1 MGAMAYPLLCLLLAQLGLG.....YMTAVGMWSLVSYDGLTIFQ 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 140259 seqs, 25548876 residues

Total number of hits satisfying chosen parameters: 140259

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3199	100.0	599	10	US-09-775-181-4
2	3119.5	97.5	1215	10	US-09-775-181-2
3	127	4.0	1078	12	US-10-002-854-2
4	124	3.9	1078	10	US-09-727-205-2
5	114	3.6	1212	12	US-10-027-923-4
6	111	3.5	1203	12	US-10-027-923-5
7	107.5	3.4	797	10	US-09-795-232-2
8	107.5	3.4	828	10	US-09-816-685-4
9	104	3.3	800	10	US-09-764-853-800
10	104	3.3	800	10	US-09-764-898-272
11	101.5	3.2	850	9	US-09-799-629-4
12	101.5	3.2	852	9	US-09-897-427A-6
13	101.5	3.2	852	10	US-09-796-338A-14
14	101	3.2	912	10	US-09-817-464-2
15	100.5	3.1	3313	10	US-09-737-149-29
16	99	3.1	669	10	US-09-801-196-28
17	98.5	3.1	530	9	US-09-858-546-2
18	98.5	3.1	763	10	US-09-819-946-4
19	98.5	3.1	841	9	US-09-897-427A-2

20	98.5	3.1	841	9	US-09-799-629-17	Sequence 17, Appl
21	98.5	3.1	841	10	US-09-819-946-2	Sequence 2, Appl
22	98	3.1	927	12	US-10-003-356-8	Sequence 8, Appl
23	97	3.0	867	10	US-09-817-464-4	Sequence 4, Appl
24	97	3.0	915	10	US-09-817-464-12	Sequence 12, Appl
25	97	3.0	922	10	US-09-817-464-14	Sequence 14, Appl
26	96	3.0	877	10	US-09-764-898-200	Sequence 200, App
27	95.5	3.0	721	10	US-09-908-322-5	Sequence 5, Appl
28	95	3.0	1624	12	US-10-090-454-2	Sequence 2, Appl
29	93	2.9	454	10	US-09-815-242-5209	Sequence 5209, Ap
30	93	2.9	810	10	US-09-976-165-34	Sequence 34, Appl
31	92.5	2.9	481	10	US-09-817-464-8	Sequence 8, Appl
32	92	2.9	358	9	US-09-974-298-86	Sequence 86, Appl
33	92	2.9	358	10	US-09-826-508-10	Sequence 10, Appl
34	92	2.9	358	10	US-09-919-172-22	Sequence 22, Appl
35	92	2.9	1165	9	US-10-026-188-2	Sequence 2, Appl
36	92	2.9	2799	9	US-10-151-736-4	Sequence 4, Appl
37	91.5	2.9	380	12	US-10-003-356-5	Sequence 5, Appl
38	91	2.8	448	9	US-10-041-016-2	Sequence 2, Appl
39	91	2.8	448	9	US-10-066-500-15	Sequence 15, Appl
40	91	2.8	448	9	US-10-174-590-408	Sequence 408, App
41	91	2.8	448	9	US-10-176-758-408	Sequence 408, App
42	91	2.8	448	9	US-10-175-737-408	Sequence 408, App
43	91	2.8	448	9	US-10-173-706-408	Sequence 408, App
44	91	2.8	448	9	US-10-175-738-408	Sequence 408, App
45	91	2.8	448	9	US-10-175-752-408	Sequence 408, App

ALIGNMENTS

RESULT 1  
US-09-775-181-4  
; Sequence 4, Application US/09775181  
; Patent No. US20020038013A1  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; TITLE OF INVENTION: No. US20020038013A1el Human Membrane Proteins and  
; TITLE OF INVENTION: Polynucleotides Encoding the Same  
; FILE REFERENCE: LEX-0129-USA  
; CURRENT APPLICATION NUMBER: US/09/775,181  
; PRIOR FILING DATE: 2001-02-01  
; PRIOR APPLICATION NUMBER: US 60/180,414  
; PRIOR FILING DATE: 2000-02-04  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 599  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-775-181-4

Query Match 100.0%; Score 3199; DB 10; Length 599;  
Best Local Similarity 100.0%; Pred. No. 1.2e-263;  
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MGAMAYPLLCLLLAQLGLGAVGASRDPOGRDPSRPTPKGKPHAQOQGRASADSSAP	60
DB	1	MGAMAYPLLCLLLAQLGLGAVGASRDPOGRDPSRPTPKGKPHAQOQGRASADSSAP	60
QY	61	WRSRTDGTTLAOKLAEEVPMVASYLYTGDHQLKRANCSGRYELAGLPCKWPALASAP	120
DB	61	WRSRTDGTTLAOKLAEEVPMVASYLYTGDHQLKRANCSGRYELAGLPCKWPALASAP	120
QY	121	SLHRALDTLTHATNFMVLMQSNKSEQNLQDDLDWYQALVMSLLGEPSSIRAAITFTST	180
DB	121	SLHRALDTLTHATNFMVLMQSNKSEQNLQDDLDWYQALVMSLLGEPSSIRAAITFTST	180
QY	181	DSLSAPAPQVFLQATREESRILLQDLSSSAPHLANATLETWFHGLRRKWRPHLHRRGPN	240
DB	181	DSLSAPAPQVFLQATREESRILLQDLSSSAPHLANATLETWFHGLRRKWRPHLHRRGPN	240

Qy	241	QGPRLGSHRRKDLGGDKSHFKWSPPYLECENGSYKPGWLVTLSAIYGLQPNLVPEP	300
Db	241	QGPRLGSHRRKDLGGDKSHFKWSPPYLECENGSYKPGWLVTLSAIYGLQPNLVPEP	300
Qy	301	RGVMKVDINLQKVDIDOCSSDGWFSGTHKCHLANSCEMPTKGLGFLVGLGAYECICKAGFYH	360
Db	301	RGVMKVDINLQKVDIDOCSSDGWFSGTHKCHLANSCEMPTKGLGFLVGLGAYECICKAGFYH	360
Qy	361	PGVLPVNNFRRRPGDQHQHISGTSKDVSEEAAYVCLPREGCPFCADDSPCFVQEDKYLRLAI	420
Db	361	PGVLPVNNFRRRPGDQHQHISGTSKDVSEEAAYVCLPREGCPFCADDSPCFVQEDKYLRLAI	420
Qy	421	ISFOGLCMLLDFVSMVLVYVYHFRKAKSIRASGLLILLETILFGSLLLFPVVILYFEPSTPR	480
Db	421	ISFOGLCMLLDFVSMVLVYVYHFRKAKSIRASGLLILLETILFGSLLLFPVVILYFEPSTPR	480
Qy	481	CILLRWARLLGFATVGTVT7LKLHRYLKVFLSRTAORIPYMTGGRYMRMLAVILLVVFVF	540
Db	481	CILLRWARLLGFATVGTVT7LKLHRYLKVFLSRTAORIPYMTGGRYMRMLAVILLVVFVF	540
Qy	541	LIGHTSSVCQNLKQIKSLIQGKTSDBHIFNMCLIDRWYMTAVGWSLSVSDGLTIFQ	599
Db	541	LIGHTSSVCQNLKQIKSLIQGKTSDBHIFNMCLIDRWYMTAVGWSLSVSDGLTIFQ	599

RESULT 2

US-09-775-181-2

; Sequence 2, Application US/09775181

; Patent No. US20020038013A1

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilbun, Erin

; TITLE OF INVENTION: No. US20020038013A1e1 Human Membrane Proteins and

; TITLE OF INVENTION: Polynucleotides Encoding the Same

; FILE REFERENCE: LEX-0129-USA

; CURRENT APPLICATION NUMBER: US/09/775,181

; CURRENT FILING DATE: 2001-02-01

; PRIOR APPLICATION NUMBER: US 60/180,414

; PRIOR FILING DATE: 2000-02-04

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 1215

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-775-181-2

```

Db      301  RGVWKVDINLQKVIDIDOCSSDGHFSGTHKCHLNNSECMPTKGLGFLVGAYECICKAGFTYH 360
Qy      361  PGVLPVNNFRRRRPQDQHSIGSTKDVSEAYVCLPCREGCPFCADDSPCFVQEDKYLRAL 420
Db      361  PGVLPVNNFRRRRPQDQHSIGSTKDVSEAYVCLPCREGCPFCADDSPCFVQEDKYLRAL 420
Qy      421  ISFOGLCMLLDFVSMLVVYHFRKAKSIRASGLLILLETILFGSLLLFPVVILYFEPSTFR 480
Db      421  ISFOGLCMLLDFVSMLVVYHFRKAKSIRASGLLILLETILFGSLLLFPVVILYFEPSTFR 480
Qy      481  CILLRWARLLGFATVYGTVTLLKLRHVLLKVLFSRTAQRIPYMTGGRVYMRMLAVILLVFWF 540
Db      481  CILLRWARLLGFATVYGTVTLLKLRHVLLKVLFSRTAQRIPYMTGGRVYMRMLAVILLVFWF 540
Qy      541  LIGHTSSVCQNLEKQISLIGQKTSOHLIFNMCLIDRWDMYMTAVG-----MWSL 589
Db      541  LIGHTSSVCQNLEKQISLIGQKTSOHLIFNMCLIDRWDMYMTAVAEFLFLNGV 594

RESULT 3
US-10-002-854-2
; Sequence 2, Application US/10002854
; Patent No. US20020132224A1
; GENERAL INFORMATION:
; APPLICANT: Scadden, David T.
; APPLICANT: Poznansky, Mark C.
; APPLICANT: Olszak, Ivona T.
; APPLICANT: Brown, Edward M.
; TITLE OF INVENTION: THE CAR RECEPTOR AS A MEDIATOR OF
; TITLE OF INVENTION: MIGRATORY CELL CHEMOTAXIS AND/OR CHEMOKINESIS
; FILE REFERENCE: M0765/7038/ERG/KA
; CURRENT APPLICATION NUMBER: US/10/002,854
; CURRENT FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/15440
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1078
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)...(1078)
; OTHER INFORMATION: Ca-sensing Receptor
US-10-002-854-2

```

[illegible]

Query Match	4.0%;	Score 127;	DB 12;	Length 1078;
Best Local Similarity	22.5%;	Pred. No. 0.026;		
Matches	71;	Conservative 48;	Mismatches 139;	Indels 58; Gaps

  

QY	311	QKVIDOCSSDGFWSGTHKC-----HLNNSE-CMPIKGLGFLVGAIECICKAG---FVHPGV	363
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
Db	475	EQVTFDEC---GDLVGNYSIINHLSPDGSI VFKEGVY----YNYAKKGERLFINEEK	527
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
QY	364	LPVNNFRRCQDQH-----ISGSTKDVSEE---AYVCLPCREG-----CPFCAD	404
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
Db	528	ILMSGFSREVPFNCSDRCUAGTRKGIIEGPTCCFCECPDGSEYSDETASACNKCPD	587
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
QY	405	D-----SPCFVQEDKYLR-----LAISFOGLMLDDFVSLVYVYHFRRAKSI	451
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
Db	588	DFWSNENHTSCIAKEIEFUSWTEPFGIALTLFAVLGIFLTAFLGVFIKERNTPIVKATN	647
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
QY	452	LILLETLFCSLLYFPVILYEPSTFERCILLRWARLLGFATVGTVTLKLRHVLKVEL	511
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
Db	648	RELSYLLFLSLLCCFSSSLFFIGEPQDWTGRLRQPAFGISFVLCISCIILVKNRVLLVFE	707
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
QY	512	SRTAQRIPYMTGGVRVMRLAV-----ILLVWFELIGWTSV-CQNLEKQISLI	559
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
Db	708	AKPTSFHRKWWGLNQFLVFLCTFMQIVICVINWLYTAPPSYRNQOELEDEIIFTCHE	767
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:
QY	560	GQKTSDDLIFNMCLI	575
	:	: : :	:
	:	: : :	:
	:	: : :	:
	:	: : :	:



Db 768 GSMALGFLIGYTCLL 783

RESULT 4

US-727-205-2  
 ; Sequence 2, Application US/09727205  
 ; Patent No. US20020064813A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ELLIS, CATHERINE E.  
 ; TITLE OF INVENTION: MONKEY CALCIUM SENSING RECEPTOR  
 ; FILE REFERENCE: GP-70665  
 ; CURRENT APPLICATION NUMBER: US/09/727,205  
 ; CURRENT FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: 60/168,342  
 ; PRIOR FILING DATE: 1999-12-01  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 1078  
 ; TYPE: PRT  
 ; ORGANISM: MACACA CYNOMOLGUS  
 09-727-205-2

Query Match 3.9%; Score 124; DB 10; Length 1078;  
 Best Local Similarity 21.7%; Pred. No. 0.046;  
 Matches 67; Conservative 49; Mismatches 129; Indels 64; Gaps 13;

QY 311 QKVIDQCSSDGFSGTHK---HLNNS-CMPIKGLGFLGAYECICKAG---FYHPGV 363  
 Db 475 EQVTDEC---GDLVGNYSIIINWHSPEDGSIVFEKVG---YNYAKKGERLFINEEK 527  
 QY 364 LPVNNFRRRPQDH-----ISGSTKDVSEE---AYVCLPCREG-----CPFCAD 404  
 Db 528 ILWSGFSREVPSNCSRDCLAGTRGIIISGPTCCFCEVCEPDGEVSDTASACNKCPD 587  
 QY 405 D-----SPCFVQEDKYLR-----LAIISFOGLCMLDFVSMVYVYHFRKAKSIRASG 451  
 Db 588 DFWSNENHTSCIAKEIEFTSWPEFGIATLFAVLGIFLTAFLVGVFIKFRNTPIVKATN 647  
 QY 452 LILLETILFGLSLLLPVVPVILYFEPSTFCILLRWALLGFATVYGTVTLKLRHVLKVP 511  
 Db 648 RELSYLLLSLCCSSSSFFIGEFQDWTCLRQPAFGISFVLCISLILVKNRVLVFE 707  
 QY 512 SRTAQIRPYMTGGRVMRLAV-----ILLVVFVFLIGWTSVVCQNKLEKQISLIGQGKT 564  
 Db 708 AKIPTSFHRKWWGLNLQFLFLVLCFTFMQIVICVIMLYTAPPSSY-RNHELE----- 757  
 QY 565 SDHLIFNMC 573  
 758 -DEIIFITC 765

RESULT 5

US-10-027-923-4  
 ; Sequence 4, Application US/10027923  
 ; Patent No. US20020142330A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Briann Gaither Bates  
 ; APPLICANT: Kamalaka Gulukota  
 ; APPLICANT: Yuhong Xie  
 ; TITLE OF INVENTION: NOVEL GLUTAMATE RECEPTOR MODULATORY PROTEINS AND  
 ; FILE REFERENCE: GNN-024  
 ; CURRENT APPLICATION NUMBER: US/10/027,923  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 60/257,589  
 ; PRIOR FILING DATE: 2000-12-22  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1212  
 ; TYPE: PRT

; ORGANISM: Homo sapiens  
 US-10-027-923-4

Query Match 3.6%; Score 114; DB 12; Length 1212;  
 Best Local Similarity 20.4%; Pred. No. 0.39;  
 Matches 80; Conservative 53; Mismatches 165; Indels 94; Gaps 16;  
 QY 219 ETEWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGGDKSHFKWSPPYLECE 273  
 Db 333 DVKWFDDYYLKLRPETNHRNP-----WFOEFWQHRFOCRLEGFPQENSKYNKT-----CN 382  
 QY 274 NG-----SYKPGWLV-TLSSAIYGL---OPNLVPEFRGV-----MKVDINLQ 311  
 Db 383 SSLTLKTHHVQDSKMGFVINAIYSMAYLHNMQMSLCPGYAGLCDAMKPIDGRKLLDSLM 442  
 QY 312 KVIDIDQCSSDGFSGTHKCHLNNSCMPKIGLGVLGAYECICKAGFYHPGVLPVNNFR 371  
 Db 443 KTNFTGVSGDTILFDENGDSGRYEIMNFKMG-----KDYFDVINVGSMWDNGEL 492  
 QY 372 RGPDQHI-----SGSTKDVSEE-----AYVCLPCREG-----CPFC 402  
 Db 493 KMDDDEVNSKSNIIIRSVCSPECEKQIKVIRKGEVSCCWTCTPCKENEVVFEYTCAC 552  
 QY 403 -----ADD-SPCFVQEDKYLR-----LAIISFOGLCMLDFVSMVYVYHFRKAKSIRA 449  
 Db 553 QLGSWPTDGLTCDLIPVQYLRWGDPEPTAAVVFACLGGLATLFTVTVFIYRDTPVVKS 612  
 QY 450 SGLILETILFGLSLLLPVVPVILYFEPSTFCILLRWALLGFATVYGTVTLKLRHVLK 509  
 Db 613 SSRELCYIILAGICLGLCTFCLIAKPKQIYCYLQRIIGLSPAMSYSALVTKTNRIARI 672  
 QY 510 FLSRTAQRI---PYMTGGRVMRLAVILLV 537  
 Db 673 -LAGSKKKICTTKPRMSACQLVITAFILICI 703

RESULT 6

US-10-027-923-5  
 ; Sequence 5, Application US/10027923  
 ; Patent No. US20020142330A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Briann Gaither Bates  
 ; APPLICANT: Kamalaka Gulukota  
 ; APPLICANT: Yuhong Xie  
 ; APPLICANT: Janet Elizabeth Paulsen  
 ; TITLE OF INVENTION: NOVEL GLUTAMATE RECEPTOR MODULATORY PROTEINS AND  
 ; FILE REFERENCE: GNN-024  
 ; CURRENT APPLICATION NUMBER: US/10/027,923  
 ; CURRENT FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 60/257,589  
 ; PRIOR FILING DATE: 2000-12-22  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 5  
 ; LENGTH: 1203  
 ; TYPE: PRT  
 ; ORGANISM: Rattus sp.  
 US-10-027-923-5

Query Match 3.5%; Score 111; DB 12; Length 1203;  
 Best Local Similarity 20.8%; Pred. No. 0.69;  
 Matches 80; Conservative 54; Mismatches 171; Indels 80; Gaps 17;  
 QY 219 ETEWFHGLRRKWRPHLHRRGNQGRGLGHSWR-----RKDGLGGDKSHFKWSPPYLECE 273  
 Db 332 DVKWFDDYYLKLRPETNLRNP-----WFOEFWQHRFOCRLEGFAQENSKYNKT-----CN 381  
 QY 274 NG-----SYKPGWLV-TLSSAIYGL---OPNLVPEFRGV-----MKVDINLQ 311  
 Db 382 SSLTLKTHHVQDSKMGFVINAIYSMAYLHNMQMSLCPGYAGLCDAMKPIDGRKLLDSLM 441  
 QY 312 KVIDIDQCSSDGFSGTHKCHLNNSCMPKIGLGVLGAYECICKAGFYHPGVLPV----- 366

```

Db 442 KTNFTGVSGDMILFDENGSPGRYETMNFKEG---KDYFDVINWGSWNGDKLMDDEV 498
QY 367 ----NNFRRRGPDQHI-SGSTKDVSEE---AYVCLPREG-----CPFC-----A 403
Db 499 WSKNNIIRSVCEPECKQIIVIRKEVSCWCTTPCKENEVDFEYTCACQLGSNPT 558
QY 404 DD-SPCFVQEDKYLRLAIISFQGLMLLDFVSMVYVYHFRKAKSIRASGLILLE 456
Db 559 DDLTGCDLIPVOYLWGDPEPTAAVFACLGLLALFVTVIIRYDTPVAVSSRELCY 618
QY 457 TILFGLSLLYFPVILYFPEPTFRICLLRWARLLGFATVYGTVTKLHRLVKVFLSRTAQ 516
Db 619 IILAGICGLYCTFCFLIAKPKQIYVQLRIGIGLSPAMYSALVTKTNRIARI-LAGSKK 677
QY 517 RI-----PYMTGGVRMLAVILLV 537
Db 678 KICTKPKRPMSCAQLVIAFILICI 702

```

```

RESULT 7
US-09-795-232-2
; Sequence 2, Application US/09795232
; Patent No. US20010012627A1
; GENERAL INFORMATION:
; APPLICANT: Anthony M. Brown
; APPLICANT: Conrad Gerald Chapman
; APPLICANT: Israel Simon Gloger
; APPLICANT: Joanne Rachel Evans
; APPLICANT: William Cairns
; APPLICANT: Hugh Jonathan Herdon
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30176-D1
; CURRENT APPLICATION NUMBER: US/09/795,232
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 09/182,728
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 9818890.7
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 797
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-795-232-2

```

```

Query Match 3.4%; Score 107.5; DB 10; Length 797;
Best Local Similarity 20.2%; Pred. No. 0.76;
Matches 113; Conservative 67; Mismatches 174; Indels 205; Gaps 30;

QY 29 QGRPSPRERTPKGPHAQOGRASADSSAPWSRSTDTGTL-----AQKLAEEVPMVAS 84
Db 23 QGHPDGP--CAPRTSPEQELPAAAAAPPVPRVPSASTGAQTFQSADARACEAERP-GVGS 79
QY 85 YLYTGDSHOLKANCGRVELAGLPGKPALASAP--SLHRLDTLTHATNFLANMLQS 142
Db 80 CKLSFPRAQASAAALRDLREACAQAS-PPPGSSGPGNALHCKIFSLRGPEGDANVSV-G 137
QY 143 NKSREQLQDDLDWYQALVMSLLEGEPSISRAAITFSDLSAPAP-QVFLQATREESRI 201
Db 138 KGTLEARNTPVVGWV-----NMSQSTVVLGTGITSVLPGSVATVATQED-- 182
QY 202 LQDLSSSAPHANATLETETFWHGLRRKWRPHLHRRGNQGRGLGHSWRRK----- 253
Db 183 -----ERGDENKARG---NWSKLDLFIILSNV 205
QY 254 ----DGLGDKSHFKWSPVPLECEN--GSKPGWLVTLSA---IYGLOPNL----- 296
Db 206 GYAVGLGN-----VWRFPYLAFOGCGGAFLLPYLWMLALAGLPFIFFLEVSLGQFASQGPV 260
QY 297 -----VPEFRG-----VNKVDINLQKVIDQC-----SSDGWFSGT 327
; FILE REFERENCE: PZJ06
; CURRENT APPLICATION NUMBER: US/09/764,853

```

```

Db 261 SVKKAIPALQCGGCIAMLIISVLIAY-VNVIICITLTYLFYLFASFVSFVLPWGSNNPWTPIE 319
QY 328 HK-----CHLNNSECMPIKGLGFLVYGAYECICKAGFYHPGVLPVNNFRRRGPDQHI 378
Db 320 CKDKTKLLDSCVSIIDHPKIQIKNSTFCMTAY-----PNVTWV-NFTSQANKTFV 368
QY 379 SGSTKDVSEEAAYVCLPREGCPFCADDDSPCFVQEDKYLRLAI---ISFOG-----LCM 428
Db 369 SG-----SEEF-----KYFVLKISAGIEYPGEIRWPLALCL 400
QY 429 LLDVFSMLVYVYHFRKAKSIRASGLLILLETILFGSLLLLYFPVY-----ILYFEFS 477
Db 401 ---FLAWITVY-ASLAKGIKTSKGVYFATPPYVVLVILLIRGVTLPCGAGAGIWF--- 453
QY 478 TPRCILLRWARLLGFATVY 496
Db 454 ----ITPKWEKLTN-ATWV 467

RESULT 8
US-09-816-685-4
; Sequence 4, Application US/09816685
; Patent No. US20020053091A1
; GENERAL INFORMATION:
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000856
; CURRENT APPLICATION NUMBER: US/09/816,685
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 828
; TYPE: PRT
; ORGANISM: Carassius auratus
US-09-816-685-4

```

```

Query Match 3.4%; Score 107.5; DB 10; Length 828;
Best Local Similarity 23.5%; Pred. No. 0.8;
Matches 61; Conservative 35; Mismatches 91; Indels 73; Gaps 12;

QY 295 NLVPEFRGVV-----KVDINL-----OKVDI--DOCSSDGFSGTHKCHLANS 335
Db 398 NTFEKEGVMTNFDANGDINLGYDVCLWDDDESEKNDIAEYYPNSSTFTTRK-NLSNI 456
QY 336 ECMPIKGLGFLVYGAYECICKAGFYHPGVLPVNNFRRRGPDQHIISGSTKDVSEAYVCLPC 395
Db 457 E-----NVLSKCSQCPCEY-----KKTAEQGHTC-----CYECLAC 489
QY 396 REG-----CPFC-----ADSPCFVQEDKYLRLAIISFQGLCMLLD 431
Db 490 AENQYSNHTDADTCSKCDTESLWSNANSKCYPKFYEFWNSGFAIALTLAALGILL- 548
QY 432 FVSMVAVVYHFRKAKSIRASGLLILLETILFGSLLLLYFPVILYFPEPTFRCLLLRWARLL 490
Db 549 LISMSALFWORNSLVVRAAGPGCHLILFSLGSFISVITFVGPESNESCVRQVIFGL 608
QY 491 GFATVYGTVTKLHRLVKVF 510
Db 609 SFTLCVSCILYKSLILLAF 628

RESULT 9
US-09-764-853-800
; Sequence 800, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PZJ06
; CURRENT APPLICATION NUMBER: US/09/764,853

```

```

QY 268 PYLECN--GSYKPGWLVTLSAIYGLQPNLVPFGRGMKVDINLQKVDIDQCSSDGWFS 325
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 609 PDQCHKNTRGGYK-----CIDLCPN-----GWTAAE-NGTCIDIDECK-----D 646

QY 326 GTHKCHLNNSCEMPIKGLGFVGLAYECICRAGFYHPGVLPVNNFRRRGPDOHISGSKDV .385
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 647 GTHQCRYNQI-CENTR-----GSYRCVCPRGYSQGV-----GRPCMDI 684

QY 386 SEEAYVCLPCREGCPFCADDSPCFQVEDKYLRRLAIISFOGLMCLLDVSMVLVYHFKAK 445
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 585 NEXEQVPKPAHQCSNTPGFSFKCICPPQOHLGLDGCKSCAGLERLPNYGTQYSYNLARFS 744

QY 446 SIR 448
      | |
Db 745 PVR 747

RESULT 11
US-09-799-629-4
; Sequence 4, Application US/09799629
; Publication No. US20030008344A1
; GENERAL INFORMATION:
; APPLICANT: ADLER, JON ELLIOT
; APPLICANT: ZOZULYA, SERGEY
; APPLICANT: LI, XIADONG
; APPLICANT: O'CONNELL, SHAWN
; APPLICANT: STASZEWSKI, LENA
; TITLE OF INVENTION: T1R TASTE RECEPTORS AND GENES ENCODING SAME
; FILE REFERENCE: 078003/0277870/RXT
; CURRENT APPLICATION NUMBER: US/09799,629
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: 60/187,546
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: 60/195,536
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/209,840
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: 60/214,213
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/226,448
; PRIOR FILING DATE: 2000-08-17
; PRIOR APPLICATION NUMBER: 60/259,227
; PRIOR FILING DATE: 2001-01-03
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 850
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-799-629-4

Query Match 3.2%; Score 101.5; DB 9; Length 850;
Best Local Similarity 20.8%; Pred. No. 2.7;
Matches 60; Conservative 47; Mismatches 107; Indels 75; Gaps 16;

QY 291 GLOPNL--VPFGRGMKVDINLQKVDIDQCSSDGWFSCTHKCHLNNSCEMPIKGLGFVLG 348
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 461 GSVPLHDVGRFNGSLRT----ERLKRWTHTSDNQ-KPVSRCSQCQCGQVRRVKGFGHSC 515

QY 349 AYECI-CKAGFYHPGVLPVNNFRRRGPDOHISGSKTDVSEAYVCLPCREGCPFCADD-- 405
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 516 CYDCVDCAGSY-----RQNP-----DIA-----CITCGQDEW 544

QY 406 SP-----CFQVEDKYLR-----LAISFOGLMCLLDVSMVLVYHFKAKSIRASG--L 452
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 545 SPERSTRCFRRSRFLAWGEPAVLLLLLLLSLAGLVAALGLGVHHRDPSLPVQASGGPL 604

QY 453 ILLETILGSLLLYFPVILYFEBSTFRCILLRWARLLGFATVYGTVTLKLRHLYKVFLS 512
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 605 ACTGLVCLG--LVCLSVLLFPQGPSARCIA---QQPLSHLPLTGCGLSTLFLQAAEIFVE 659

QY 513 RTAQRIPYMTGGRVWRML-----AVILLVVF-----WFLIGWTSV 548

```



RESULT 15  
US-09-737-149-29  
; Sequence 29, Application US/09737149  
; Patent No. US20020077466A1  
; GENERAL INFORMATION:  
; APPLICANT: Spaderna, Steven K  
; APPLICANT: Quinn, Kerry E.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Muralidhara, Padigaru  
; APPLICANT: Spytek, Kimberly A.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 15966-620 CIP  
; CURRENT APPLICATION NUMBER: US/09/737,149  
; CURRENT FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: 60/170,564  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: 60/173,165  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,362  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 60/173,544  
; PRIOR FILING DATE: 1999-12-29  
; PRIOR APPLICATION NUMBER: 60/174,404  
; PRIOR FILING DATE: 2000-01-04  
; PRIOR APPLICATION NUMBER: 60/174,962  
; PRIOR FILING DATE: 2000-01-07  
; PRIOR APPLICATION NUMBER: 60/223,929  
; PRIOR FILING DATE: 2000-08-09  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 29  
; LENGTH: 3313  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-09-737-149-29

Query Match 3.1%; Score 100.5; DB 10; Length 3313;  
Best Local Similarity 21.2%; Pred. No. 23;  
Matches 84; Conservative 53; Mismatches 143; Indels 117; Gaps 21;

QY 81 DVASYLY---TGDSHQKRAN-CSGYELA-GLPKWFPALASAPSLHRALDTLTHATN 134  
DB 1192 DVSDHLFYSEFGRNELQLLVVNOTSGELSRKLDNNRPLVASMLVTVDGLHSVT-AQC 1250  
QY 135 FLNVMLOSNNKREONLODDLD--WYQ---ALVWSLLEGEPSISRAAITFTSDLSAPAP 188  
DB 1251 VLRVVITTELLANSLTVLENWQERFLSPLGLHGLEGAIV-----LATPTE 1299  
DB 189 QVFLOATREES----RILQDLSSAPHLANATLETWFHGLRRKWRPHLHRRGPNQGP 244  
DB 1300 DVFIQNDTDVGGTVLVNVSFALAPRGAGAGAPWESS--EELQEQLYVVRALAA 1357  
QY 245 GLGHSWRKRDGLGGDKS-----HFKWSPYLECENGSKPGWLVTLSS 287  
DB 1358 SL-----LDVLFPDDNVGLPCENYMKCVSVLRPDSSAPFLASASTLFRP-----IQ 1405  
QY 288 AIYGLQPNLVPEFRGVKVDINLQKVIDQCSDGWFSCETHKCHLNNSCEMPLIKGLGV 347  
DB 1406 PTAGLCRCRPPGFTG-----DFCETELDLGYS-----NPCR-NGGACARRE----- 1445  
QY 348 GAYECICKAGFYHPGVLVFNFRFRGPDQHISGSTKDVSEAYVCLP--CREG----- 398  
DB 1446 GGYTCVCRPRF-----TGCECELDEAGRCVPGVCRNGGCTCTNAP 1485  
QY 399 -----CPF-CADDSF-CFVQEDKYLRLAITSFQGL 426  
DB 1486 NGGFRQCPCAGGAFEGPRCEVAARSPPSSFVNFRL 1522

**THIS PAGE BLANK (USPTO)**